Copyrighted Material

Third Edition

FINANCIAL TRADING AND INVESTING



John L. Teall

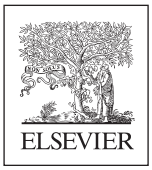
Copyrighted Material

FINANCIAL TRADING AND INVESTING

FINANCIAL TRADING AND INVESTING

THIRD EDITION

John L. Teall
Johns Hopkins University





ACADEMIC PRESS
An imprint of Elsevier

Contents

Preface xi
Acknowledgments xiii

1. Introduction to Securities Trading and Markets

1.1 Trades, Traders, Securities, and Markets 1
1.2 Securities Trading 4
1.3 Bargaining 6
1.4 Auctions 7
1.5 Introduction to Market Microstructure 12
1.6 Orders, Liquidity and Depth 14
1.7 Day Trading 17
1.8 Trading and the Broader Economy 24
Additional Reading 29
References 29
1.9 Exercises 30

2. Securities Markets

2.1 Exchanges and Floor Markets 35
2.2 The Way It Was 36
2.3 Over-the-Counter Markets and Alternative Trading Systems 45
2.4 The Decline of Brick and Mortar 49
2.5 Crossing Networks and Upstairs Markets 53
2.6 Fixed-Income Securities and Money Markets 55
2.7 Markets Around the World 64
2.8 Currency Exchange and Markets 66
2.9 Cryptocurrency and Bitcoin Trading 72
2.10 Commodities and Futures 74
2.11 Swaps and Swap Execution Facilities 76
Additional Reading 79
References 79
2.12 Exercises 80

3. Financial Market Utilities and Securities Trading Support

3.1 Securities Trading Infrastructure 83
3.2 Price and Quotation Systems 84
3.3 Brokerage Operations 85
3.4 Clearing and Settlement 87
3.5 Other Financial Market Utilities and Functions 94
3.6 Systemically Important Financial Market Utilities 96
Additional Reading 98
References 98
3.7 Exercises 99

4. Institutional Trading

4.1 Institutions and Market Impact 101
4.2 Registered Investment Companies 102
4.3 Unregistered Investment Companies 106
4.4 Best Execution, Execution Costs, and Price Improvement 108
4.5 Algorithmic Trading 113
4.6 Dark Pools 115
4.7 Stealth and Sunshine Trading 117
4.8 High-Frequency Trading 118
4.9 Flash Trading and Sponsored Access 125
Additional Reading 128
References 128
4.10 Exercises 129

5. Regulation of Trading and Securities Markets

5.1 Regulatory Approaches and the Regulatory Balance 133
5.2 Pre-1930s Securities Regulation: The Background 136

5.3 U.S. Securities Market Legislation: The Foundation 139
5.4 Crises and Updating the Regulatory System 142
5.5 Deregulation, Corporate Scandals, and the Financial Crisis of 2008 146
5.6 Dodd–Frank 149
5.7 Government Oversight of Self-Regulation 151
5.8 Impact of Regulatory Activity 155
5.9 Regulation: The International Arena 159
5.10 Privatization of Regulation and Exchange Rules 162
Additional Reading 165
References 165
5.11 Exercises 167

6. Adverse Selection, Trading, and Liquidity

6.1 Market Microstructure, Liquidity, and Spreads 169
6.2 Information and Trading 171
6.3 Noise Traders 172
6.4 Adverse Selection in Dealer Markets 173
6.5 Adverse Selection, Dealer Inventories, and the Spread 177
6.6 Risk, Risk Aversion and Dealer Spreads 180
Additional Reading 183
References 183
6.7 Exercises 184
6.A.1 The Kyle Adverse Selection Model 188

7. Random Walks, Risk, and Arbitrage

7.1 Market Efficiency and Random Walks 195
7.2 Risk 198
7.3 Arbitrage 204
7.4 Limits to Arbitrage 206
Additional Reading 209
References 209
7.5 Exercises 210
7.A.1 Return and Risk Spreadsheet Applications 215
7.A.2 A Primer on Black-Scholes Option Pricing 215
7.A.3 Estimating Implied Black-Scholes Variances 221

8. Arbitrage and Hedging With Fixed Income Instruments and Currencies

8.1 Arbitrage With Riskless Bonds 227
8.2 Fixed Income Hedging 229
8.3 Fixed Income Portfolio Immunization 233
8.4 Term Structure, Interest Rate Contracts, and Hedging 239
8.5 Arbitrage With Currencies 248
8.6 Arbitrage and Hedging With Currency Forward Contracts 250
8.7 Hedging Exchange Exposure 255
Additional Reading 257
References 257
8.8 Exercises 257

9. Arbitrage and Hedging With Options

9.1 Derivative Securities Markets and Hedging 261
9.2 Put–Call Parity 263
9.3 Options and Hedging in a Binomial Environment 264
9.4 The Greeks and Hedging in a Black-Scholes Environment 274
9.5 Exchange Options 282
9.6 Hedging Exchange Exposure With Currency Options 284
Additional Reading 285
References 286
9.7 Exercises 286
9.A.1 The Binomial Model: Additional Considerations 290
9.A.2 Deriving the Black-Scholes Model 293

10. Evaluating Trading Strategies and Performance

10.1 Evaluating Investment Portfolio Performance 297
10.2 Market Timing Versus Selection 303
10.3 Trade Evaluation and Volume-Weighted Average Price 306
10.4 Implementation Shortfall 309
10.5 Value at Risk 311
Additional Reading 315
References 315
10.6 Exercises 316

11. The Mind of the Investor

11.1 Rational Investor Paradigms 321
11.2 Prospect Theory 325
11.3 Behavioral Finance 329
11.4 Body and Mind of the Trader 339
11.5 Neurofinance: Getting Into the Investor's Head 344
11.6 The Consensus Opinion: Stupid Investors, Smart Markets? 346
Additional Reading 350
References 350
11.7 Exercises 354

12. Market Efficiency

12.1 Introduction to Market Efficiency 359
12.2 Weak Form Efficiency and Technical Analysis 360
12.3 Testing Momentum and Mean Reversion Strategies 370
12.4 Semistrong Form Efficiency 373
12.5 The Event Study Methodology 379
12.6 Strong Form Efficiency and Insider Trading 389
12.7 Anomalous Efficiency and Prediction Markets 391
12.8 Epilogue 394
Additional Reading 395
References 395
12.9 Exercises 399

13. Trading Gone Awry

13.1 The Dark Side of Finance 403
13.2 Illegal Insider Trading 404
13.3 Front Running and Late Trading 411
13.4 Bluffing and Market Manipulation 413
13.5 Payment for Order Flow 423
13.6 Fat Fingers, Hot Potatoes, and Technical Glitches 425
13.7 Rogue Trading and Rogue Traders 427
13.8 Hiding Trading Losses with Ponzi Schemes 432
Additional Reading 435
References 435
13.9 Exercises 437

Mathematics Appendix 439
End-of-Chapter Exercise Solutions 457
Glossary 519
Index 539

Academic Press is an imprint of Elsevier
125 London Wall, London EC2Y 5AS, United Kingdom
525 B Street, Suite 1800, San Diego, CA 92101-4495, United States
50 Hampshire Street, 5th Floor, Cambridge, MA 02139, United States
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, United Kingdom

Copyright © 2023 Elsevier Inc. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Details on how to seek permission, further information about the Publisher's permissions policies and our arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: www.elsevier.com/permissions.

This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

Notices

Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

ISBN: 978-0-323-90955-6

For Information on all Academic Press publications
visit our website at https://www.elsevier.com/books-and-journals

Publisher: Katey Birtcher
Editorial Project Manager: Chris Hockaday
Production Project Manager: Gayathri S
Cover Designer: Miles Hitchen

Typeset by MPS Limited, Chennai, India
Printed in the United States of America
Last digit is the print number: 9 8 7 6 5 4 3 2 1



To Emily:

Artist, Scholar, Humanitarian

Preface

There is an apocryphal story told in certain chess circles about a teenage boy that simultaneously challenged the great Russian chess champion Alexander Alekhine and the Russo-German Grandmaster Efim Bogoljubov. The boy had separately approached each of the two grandmasters to play games of correspondence chess. The two champions, without knowledge of the other's participation, accepted the boy's offers to play for money, with odds favorable to the boy. The chess greats, who happened to be friends, discovered the scheme in a much later conversation, and realized that the boy was merely acting as an intermediary between them, relaying moves between the two grandmasters. The boy, because of his favorable odds, was assured of netting a profit on the intermediation regardless of who won the match.

Intermediation is the keystone of the world financial system. While much essential financial decision-making focuses on the time and risk dimensions of investing, intermediation is the process that actually moves capital from businesses, institutions, and individuals with surplus funds to those that need them. Trading is central to this intermediation. Some traders have great expertise in valuation, quantitative analysis, or other investments tools, and serve important roles in this intermediation process. Other traders merely serve as intermediaries, quoting their prices based on the trading moves played by the experts. Their roles and opportunities are considerable as well. Regardless, all traders need to understand their markets and manage their risks. This is the objective to which *Financial Trading and Investing* strives.

Securities trading is a very broad topic, and merits far more discussion than could ever fit into a single volume. There are thousands of books on trading, all-too-many of which present their authors' opinions on "how to trade your way to riches." There are quite a few outstanding books that focus on fairly narrow aspects of trading, such as market microstructure, algo trading, risk management or day-trading, etc. *Financial Trading and Investing* seeks to provide a broad overview to trading and investing, much as any standard investments text. However, the focus of this text is on trading, trading institutions, markets, and the institutions that participate in, facilitate, and regulate trading activities. This book is not intended to replace a standard investments text. It does not introduce important general investment concepts such as portfolio theory and fundamental analysis; it focuses on trading. However, material on these extra topics is available on the text web site at https://www.elsevier.com/books-and-journals/book-companion/9780323909556 and https://educate.elsevier.com/9780323909556.

Financial Trading and Investing is intended to serve as a primary textbook for a university course focusing on trading. Most but not all students in a trading course will have already had some exposure to introductory finance or investments courses. The book should also be useful in a number of other courses, including those dealing with general investments,

securities law, and classes based on student managed funds, in some cases, perhaps as a secondary text. The text discusses recent, important, newsworthy, and controversial trading tactics such as algo trading, dark liquidity pools, late trading, co-location, high frequency trading, and flash orders. It explains a number of important valuation and hedging techniques. It introduces some of the most recent market and technology advances and innovations such as Designated Market Makers (DMMs), Supplemental Liquidity Providers (SLPs), SuperMontage, ATSs, and ECNs. The book explains the "plumbing" of securities markets, including clearing and settlement, along with central counterparties.

When I began teaching courses in trading years ago at Rensselaer Polytechnic Institute, enrollment in my classes was quite diverse. Students included those with typical undergraduate and MBA-level business studies backgrounds, with one or a few finance courses behind them (sometimes, but not always including one in investments), as well as a single course in statistics (typically) and perhaps calculus. Within a few years, students majoring in financial engineering or quantitative finance began to outnumber other students. These students had stronger quantitative backgrounds, but often grasped few institutional details and sometimes required extra patience due to their verbal skills. The most difficult students to accommodate were often doctoral students in technical programs, including engineering, physics, math, and computer science. These students were superb, but lacked any academic background in business or finance, and had very serious reasons for enrolling in my courses. I found that the best way to accommodate such a broad array of students was to assume that they had limited or no academic backgrounds in finance and might need refreshers in statistics, linear algebra, and calculus, though the mathematics refreshers were clearly not necessary for doctoral students. Regardless, I wrote the book to accommodate students with limited academic backgrounds, but who were seriously interested in trading.

While coverage in my own trading course is sequenced by the ordering of the chapters, the text is designed to be very flexible. The sequence of course coverage need not correspond with chapter ordering in the text. Most chapters in this text can be presented without significant reliance on preceding chapters. For example, many instructors would prefer their class readings in behavioral finance (Chapter 11) and market efficiency (Chapter 12) to take place early in their courses. Practically any ordering of chapters is possible, though Chapter 1 is the most logical starting point and readers unfamiliar with Black-Scholes should familiarize themselves with the Appendix to Chapter 7 before proceeding to the remainder of Chapter 7 and to Chapter 9. In addition, readers might feel more comfortable reading Chapter 2 before Chapters 3 and 4, and Chapters 4 and 5 before Chapter 13. In general, the text is organized as follows:

1. Chapters 1 through 6: Introduction to trading, trading mechanisms, clearing and settlement, regulation, and market microstructure
2. Chapters 7 through 10: Risk, arbitrage, hedging, risk management, and performance evaluation
3. Chapters 11 through 13: Behavioral aspects of trading and investing, market efficiency, and market abuses and failure

Many trading courses implement a software or internet-based trading simulation such as FTS, Rotman Interactive Trader, or TraderEx. This book was designed for use as a primary text in such learning environments. In fact, integration of simulation packages into

the trading course has played a key role in the development of this book, and close to a third of class time in my own course is dedicated to simulations. The book has ancillary materials to aid in the presentation and analysis of cases and demonstrations from these simulations.

In addition to the simulation-based ancillary materials, which are located at the book's student companion site at https://www.elsevier.com/books-and-journals/book-companion/9780323909556 and instructor resources at https://educate.elsevier.com/9780323909556, the text has several other pedagogical features. First, each chapter provides a number of exercises, to which detailed solutions are offered at the end of the text. There are a number of appendices, both to chapters and at the end of the text. For example, text appendices review basic statistics and mathematics concepts that are most useful to traders, particularly in the measurement and management of risk. The web-based ancillary materials include additional questions and problems that can be used for in-class discussions or exams. Additional discussion on the presentations of concepts and integration with simulation packages are offered there as well.

I welcome your comments and suggestions regarding this book. I can be contacted by e-mail at jteall@jteall.com or by using contact information available on my web page at http://www.jteall.com.

ACKNOWLEDGMENTS

I am fortunate to have taught a number of excellent students who have offered many suggestions and corrections on chapter drafts for this book and its second edition. Several colleagues, friends, and reviewers made helpful contributions, including T.J. Wu, Tina Prodromou, Peter Knopf, Anthony Joseph, and Iman Palit. I am especially grateful to Lynda S. Livingston and Vassilios Gargalas, who provided many particularly helpful comments and ideas. Stan Wakefield and editors and staff at Elsevier, including Scott Bentley, Gayathri S, Anusha Sambamoorthy, Katey Birtcher and Chris Hockaday, were essential in bringing the book and this third edition to fruition. In addition, a number of anonymous reviewers were most helpful. Of course, my lovely wife Anne, who no longer requires that I work one day a week, and Emily, who had been such a delightfully pleasant adolescent, and now lives in my basement, always boost my productivity and morale.

CHAPTER

1

Introduction to Securities Trading and Markets

1.1 TRADES, TRADERS, SECURITIES, AND MARKETS

A *trade* is a security transaction that creates or alters a portfolio position based on an investment decision. The trade action and the trade decision follow the investing decision to buy, sell, or otherwise take a position in an asset or instrument. The trade decision concerns how to execute the investment decision, in which markets, at what prices and times and through which agents. Trade decisions are more concerned with the speed, costs, and risks associated with executing the transaction, while investing emphasizes selection of the security. The investment strategy is linked to the rationale to buy or to sell (e.g., to exploit undervaluation). Investing is a positive sum game, allocating capital towards production and risk management. Trading is often a zero sum game, except to the extent that it facilitates investing, produces information, and improves markets by enhancing liquidity (ease of buying and selling securities) and reducing execution costs. While investing and trading activities might be distinguished from one another, there is a significant overlap between them, where high-turnover short-term investing, also referred to as trading, blurs the distinction between the two.

Traders compete to generate profits, seeking compatible counterparties in trade and seeking superior order placement and timing. *Proprietary traders* seek profits by trading on their own accounts while *agency traders* trade as commission brokers on behalf of clients. Most proprietary traders are *speculators* who focus on profits derived from price changes, *arbitrageurs* who focus on price discrepancies, and *hedgers* who seek to control risk. *Dealers*, who trade directly with clients, and *brokers*, who seek trade counterparties for clients, facilitate the trading process. Brokers act as agents for investors, buying and selling for them on a commission or mark-up basis. Dealers maintain *quotes* (*bids*, which are solicitations to purchase and *offers*, which are solicitations to sell) and buy and sell on a profit basis. The typical quote specifies the security, the proposed transaction's price, and the number of units to be exchanged at this price. The spread is simply the difference between the best offer and best bid. *Buy side traders* such as individual investors, mutual funds, and pension

DOI: https://doi.org/10.1016/B978-0-323-90955-6.00001-X

© 2023 Elsevier Inc. All rights reserved.

funds buy exchange or liquidity services. This means that buy side traders make investment decisions to take positions in securities and seek counterparties to buy from or sell to. *Sell side traders* such as day traders, market makers, and brokers provide liquidity and markets to buy side traders. Sell side traders stand by awaiting orders from buy side traders. One should not confuse buy side and sell side traders with buyers and sellers of securities; buy-side traders buy liquidity services and sell-side traders sell liquidity services. Regulatory definitions for traders will be discussed shortly.

Trading Illustration

Consider a hypothetical scenario where an equities analyst working for the Flagellan Fund has recommended that the fund purchase shares of the General Engine Company. The fund's portfolio manager agreed, and placed an order for 50,000 shares with the fund's in-house trader. The in-house trader decided to transact 10,000 shares herself, and placed buy orders of 20,000 shares each with two brokers, Kanteven-Fitzgerald and Themus Trading. Counterparties for the trades will be found in various trading venues, including exchanges and alternative trading systems. Flagellan is considered a buy side market participant, not because it is buying shares, but because it is buying liquidity. Its counterparties in trade will probably be a combination of buy side and sell side market participants.

Securities and Instruments

The term "security" has several alternative definitions. In this text, we generally define a security to be a tradable claim on the assets of an institution or individual. Where real assets contribute to the productive capacity of the economy, securities are financial assets that merely represent claims on real assets. Many securities represent ownership (e.g., shares of stock) or creditorship (e.g., bond) of an institution (e.g., corporation). Some instruments will represent obligations to buy or sell (e.g., futures contracts) or options to buy or sell. Most corporate securities imply either fixed claims (such as bonds that typically involve fixed interest and principal repayments) or residual claims (such as common stock, whose owners receive assets remaining after creditors' and other claims have been satisfied). Most securities are marketable to the general public, meaning that they can be sold or assigned to other investors in the open marketplace.

In a 1946 decision (SEC v. Howey Co., 328 U.S. 293), the U.S. Supreme Court defined a security to be an investment contract based on the following four criteria, which are collectively referred to as the *Howey Test*:

1. The involvement of an investment of money or other valued assets.
2. The formation of and investment into a common enterprise.
3. An expectation of profits from this investment.
4. The use of a third party to promote the offering.

Some of the more common types of securities and tradable instruments are classified and briefly introduced in the following:

1. *Debt securities*: Denote creditorship of an individual, firm, or other institution. They typically involve payments of a fixed series of payments to interest or amounts towards principal along with principal repayment. Examples include:
   Bonds: Long-term debt securities issued by corporations, governments, or other institutions.
   Treasury securities: Debt securities issued by the U.S. Treasury of the federal government.
2. *Equity securities (stock):* Denote ownership in a business or corporation. They typically permit for dividend payments if the firm's debt obligations have been satisfied. The two primary types of marketable equity issued by corporations are:
   Common stock: Security held by the residual claimant or owner of the firm
   Preferred stock: Stock that is given priority over common stock in the payment of dividends and liquidation; preferred stock holders must receive their dividends if common stock holders are to be paid dividends
3. *Derivative securities:* Have payoff functions derived from the values of other securities, rates, or indices. Some of the more common derivative securities are:
   Options: Securities that grant their owners rights to buy or sell an asset at a specific price on or before the expiration date of the security. Options on stock are among the most frequently traded. The two types of stock options are[1]:
      Call: A security or contract granting its owner the right to purchase a given asset at a specified price on or before the expiration date of the contract
      Put: A security or contract granting its owner the right to sell a given asset at a specified price on or before the expiration date of the contract
   Futures contracts: Instruments that oblige their participants to either purchase or sell a given asset at a specified price on the future settlement date of that contract. We will discuss later the differences between futures and options contracts. Investors may take either a long or a short position in a futures contract. A *long* position obligates the investor to purchase the given asset on the settlement date of the contract and a *short* position obligates the investor to sell the given asset on the settlement date of the contract.
   Swaps: Provide for the exchange of cash flows associated with one asset, rate, or index for the cash flows associated with another asset, rate, or index.
4. *Commodities*: Contracts, including futures and options on physical commodities such as oil, metals, corn, and so on. A commodity contract will be characterized by five features: the commodity (e.g., heating oil no. 2), the exchange on which the contract is traded (e.g., NYM for New York Mercantile Exchange), the size of the contract (e.g., 42,000 gallons), the settlement price per unit (e.g., $2.75 per gallon), and the settlement date or delivery month (e.g., September 2010).
5. *Currencies*: For transactions to be executed between countries with different currencies, some type of foreign exchange (also called currency exchange, FOREX or FX) must take place. Foreign exchange is simply the trading of one currency for another. Exchange rates denote the number of units of one currency that must be given up for one unit of a second

[1]Actually, these are "plain vanilla options." There are also a large number of exotic options with more complex contractual terms.

currency. Exchange transactions can occur in either *spot* or *forward* markets. In the spot market, the exchange of one currency for another occurs when the agreement is made. For example, dollars may be exchanged for euro now in an agreement made now. This would be a spot market transaction. In a forward market transaction, the actual exchange of one currency for another actually occurs at a date later than that of the agreement. Thus, traders could agree now on an exchange rate for two currencies at a later date. Spot and forward contract participants take one position in each of two currencies:

a. Long: An investor has a "long" position in the currency that he will accept at the later date.

b. Short: An investor has a "short" position in the currency that he must deliver in the exchange.

Currency futures contracts are much like forward contracts, standardized to trade on exchanges with specified settlement prices, dates, and contract sizes. Futures contracts normally provide for margin and marking to the market.

6. *Indices*: Contracts pegged to measures of market performance such as the Dow Jones Industrials Average or the S&P 500 Index. These are frequently futures contracts on portfolios structured to perform exactly as the indices for which they are named. Index traders also trade options on these futures contracts.

It is important to note that this list of security types is far from complete; it only reflects some of the instruments most frequently discussed in this book. In addition, most of the instrument types will have many different variations.

Investors purchase securities for investment purposes. Their purchases may be motivated by profit motives or by risk management. As we discussed earlier, profits might be obtained through successful *speculating* (forecasting correctly the direction of security prices and payments associated with securities). Speculators are highly dependent on quality information. A second potential source of profits is *arbitrage*, the simultaneous purchase and sale of the same or substantially similar asset at different prices. Arbitrage succeeds when the sum paid for purchased assets is less than the sales proceeds of sold assets. Arbitrageurs are highly dependent on low transaction costs and speed of trade execution. Profits can also be obtained through market making, that is, by providing liquidity while profiting from differences between bid and offer prices. Many investors will profit from all three types of trading motivations. Investors might also purchase securities to manage risk. For example, some securities serve as excellent hedges for others. Securities are purchased from selling *counterparties* and the acquisition and liquidation processes are referred to as trading.

We will regularly discuss margin in its various forms in later chapters. In brief, the term *margin* refers to funds posted at the initiation of an investment position by a trader to ensure the trader fulfills her trading obligations. Margin can also refer to borrowing money to invest in securities—that is, for example, to buy stock on margin.

1.2 SECURITIES TRADING

Trading occurs in securities markets, physical or virtual, where traders communicate with one another and execute transactions. The basic function of a market is to bring

together buyers and sellers. Most markets also provide information in the price discovery process, and the information revealed in this process is a function of market structure. This book will focus on the process of trading and design of markets, and will discuss investing and certain investment decisions as well when appropriate.

The first component of a trade involves the *acquisition of information* and quotes. Quality information and transparency are crucial to price discovery, the process by which the market discovers prices. Traders want to know prices at which they can buy (ask or offer prices) and sell (bid prices), along with the more recent prices of actual executions (e.g., last). Real-time and time-delayed quote and price services can be purchased from data vendors such as Bloomberg, Refinitiv Eikon, and Factset or obtained through a variety of broker, media and other sources.[2] *Transparent* markets disseminate high-quality information quickly to the public. Generally, markets that communicate real-time bids, offers, and executed trade information are more transparent than dealer markets that do not. *Opaque markets* are those that lack transparency.

The second component of the trade is the *routing* of the trade order. Routing involves selecting the broker(s) to handle the trade(s), deciding which market(s) will execute the trade(s) and transmitting the trade(s) to the market(s). An individual investor might route her trade through her broker who might route it to a major exchange. The third component of the trade is its *execution*, when the security is actually bought or sold. Buys are matched and executed against sells according to the rules of that market. Rules might provide for electronic execution, telephone execution, or direct human (face-to-face) execution. Finally, the fourth component of the trade is its *confirmation*, *clearance*, and *settlement*. Clearance is the recording and comparison of the trade records, and settlement involves the actual delivery of the security and its payment. *Trade allocation* might also be a part of this final process, as large institutional orders involving many clients and many transactions are allocated to various clients. Settlement is the actual transfer of securities and payments.

Algorithmic Trading: A Brief Introduction

Large orders move markets. For example, a large buy order increases security prices, which, obviously, is not good for the buyer placing the order. Typical transaction sizes in many markets such as the NYSE used to be very small relative to the order sizes placed by institutional traders. This meant that large institutional trades could have a significant and unwanted impact on execution prices unless larger orders were broken into smaller orders, to disguise them as uninformed small investor orders. *Algorithmic trading* refers to automated trading with the use of computer programs for automatically submitting and allocating trade orders among markets and brokers as well as over time so as to minimize the price impact of large trades. Many institutional and agency traders employ algorithmic trading (also called automated trading, black box trading, and robotrading) to break up large orders into smaller orders to reduce execution risk, preserve anonymity, and minimize the price impact of a trade. Significant portions of these orders might be withheld from public display to minimize their price impact in the market. The hidden portions of these large institutional orders are sometimes referred to as *dark liquidity pools* because they are hidden

[2]More details on sources of trade information will be provided later in this chapter.

from the public. Orders are often partially revealed, in which case they are called *iceberg* or *hidden-size* orders, with brokers instructed not to reveal the full size of the order. Rooted in program trading (defined by the NYSE as computer-initiated trades involving 15 or more stocks with value totaling more than $1 million) dating back to the 1980s, algorithmic trading accounts for between 30% and 80% of trade volume in many markets.

Strategies used by algorithms vary widely. Generally, algorithmic trading results from mathematical models that analyze every quote and trade in the relevant market, identify liquidity opportunities, and use this information to make intelligent trading decisions. Some algorithmic trading models seek to trade at or better than the average price over a day (e.g., volume weighted average price—VWAP, pronounced Vee-whap) and others seek to execute slowly so as to have minimal price impact. Algorithms sometimes are set to produce more volume at market opens and closes when trading volume is high, and less during slower periods such as around lunch. They can seek to exploit any arbitrage opportunities or price spreads between correlated securities. In addition, algorithmic trading can also account and adjust for trading costs such as commissions and taxes as well as regulatory issues such as those associated with short selling (borrowing stock, then selling it). However, algorithmic trading does have risks, such as leaks that might arise from competitor efforts to reverse engineer them. Many algorithms lack the capacity to handle or respond to exceptional or rare events. This can make algorithmic trading very risky. In addition, any malfunction, including a simple lapse in communication lines, can cause the system to fail. Thus, careful human supervision of algorithmic trading and other safeguards are crucial.

While algorithmic trading traditionally refers to the trade slicing and execution methods discussed above, the term is sometimes used in a more general sense to include "alpha models," which are used to make trade decisions to generate trading profits or control risk. Thus, more generally, algorithmic trading could be defined as trading based on the use of computer programs and sophisticated trading analytics to execute orders according to predefined strategies. Thus, algorithmic trading can be generalized to include program trading as defined by the NYSE earlier in this section. Regardless, algorithmic trading is highly dependent on the most sophisticated technology and analytics.

1.3 BARGAINING

One of the most important aspects of the trading process is to arrive at a price agreeable to all parties of the trade. *Bargaining* is the negotiation process over contract terms that occurs between a single buyer and a single seller for a single transaction. The negotiation can be over price, quantity, security attributes, or any other factor important to one or both counterparties. More generally, bargaining occurs when two or more economic agents have in common an interest to cooperate, but have conflicting interests over the details of cooperation. Bargaining is the process used by agents to seek an agreement.

Bargaining is useful in the trading of securities when transaction sizes are large enough for the benefits of negotiation to exceed the costs of negotiation. Bargaining on prices occurs between floor brokers on the NYSE when the transaction size is large enough to justify the cost of personal interaction. *Liquidnet*, an "upstairs" market that matches institutional buyers and sellers of large blocks of equity securities and other instruments, enables

institutions to directly bargain and trade confidentially with one another. Liquidnet provides an electronic format for this negotiation process. Traders use the system by entering symbols for securities that they wish to buy or sell. If Liquidnet sees a potential match, it notifies the counterparties, who bargain anonymously with each other in a virtual meeting room using the Liquidnet system. In an effort to protect trader confidentiality, Liquidnet rates traders with respect to the likelihood that they will actually execute transactions that Liquidnet arranges. Traders seeking to avoid front-running and quote-matching can request that trade information be hidden from other traders who are deemed unlikely to actually successfully negotiate transactions.[3] When matched, trade counterparties negotiate directly with each other the terms of their trade. Counterparties will know prices from other markets, leaving relatively little room for significant pricing disagreements. If they come to terms, they report the transaction to Liquidnet, which arranges the transfer and takes a commission on each share. In January 2017, Liquidnet claimed that the average size of its institutional orders was approximately 42,000 shares, compared to average order sizes of less than 300 shares in the NYSE and NASDAQ markets. In the NYSE and NASDAQ markets, institutions seek to execute large orders with small counterparties, requiring them to break up orders and work them over time. It is very difficult for institutions to disguise their intents with respect to order directions and sizes, placing them at obvious disadvantage in the trading arena. Obviously, trading against smaller orders is not the most effective means of satisfying institutional liquidity needs. Bargaining is often an expensive means to discover a price, but its costs may be less than its benefits for large transactions.

Bargaining power is the relative ability of one competitor to exert influence over another. In a trading scenario, bargaining power typically reflects one trader's ability to influence the transaction price. Relative bargaining power among traders is typically a product of the following:

- Patience and liquidity, which increases bargaining power
- Risk aversion, which reduces bargaining power
- Credible alternatives and options that enhance bargaining power
- The cost of backing down, which decreases bargaining power
- Superior information, which increases bargaining power
- Reputation with respect to strength, staying power, and resolve, all of which enhance bargaining power

1.4 AUCTIONS

An *auction* is a competitive market process involving multiple buyers, multiple sellers, or both. An auction is the process of trading a security through bidding, and then placing

[3] An example of front-running is when a broker with information about a client's order uses this information to set a bid or offer in her own order so as to exploit the client's order. Quote matching occurs when a small trader places an order one tick from that of a large trader so as to profit from the large trader's transaction price pressure, or to use the large trader as a counterparty should prices reverse. Both types of these disruptive transactions will be discussed in Chapter 13.

it to the winning bidder.[4] Vickrey (1961) demonstrated that optimal bids are increasing in bidders' values; therefore, the auctioned object will be won by the bidder who values it the most. Auctions are a useful and cost-effective method for pricing a security with an unknown value. That is, auctions are useful price discovery processes.

A *Walrasian auction* is a simultaneous auction where each buyer submits to the auctioneer his demand and each seller submits her supply for a given security at every possible price. Theoretically, the Walrasian auctioneer conducts a series of preliminary auctions called *tâtonnement* to determine supply and demand levels at various prices. In the actual auction, supply is balanced against demand meaning that the resulting transaction price is set so that the total demand from all buyers equals the total supply from all sellers. Thus, the Walrasian auction finds the clearing price that perfectly matches the supply and the demand. While the Walrasian auction has been used primarily as a theoretical construct, it does closely resemble certain financial auctions, such as that conducted in the London Bullion Market Association fixing process.[5] In this market, five representatives of firms would meet twice each day at the offices of N.M. Rothschild where the chairman called out prices so that firm representatives could announce their levels of supply and demand at announced prices. Until the imbalance between supply and demand cleared, new prices were announced. Many markets use similar procedures to open trading for the day.

An *English auction* or ascending bid auction, used by the English art and antique auctioneers, is typically employed to sell automobiles, farm equipment, and animals as well as art. English auction participants bid openly against one another, with each successive bid higher than the previous one. Because the auction is open and involves public sequences of bids, it provides for some degree of price discovery before it concludes. The auction ends and a "winner" is declared when no participant is willing to bid higher.

The traditional *Dutch auction* or descending bid auction, used in the 17th-century Dutch tulip bulb auctions, begins with the auctioneer calling out a high offer price, which is reduced until some participant submits the first and highest bid, crying out "mine." The "winner" pays that price. Thus, the Dutch auction does reveal some information concerning a price ceiling on the auctioned object. The Dutch auction is particularly useful for matching a number of identical goods to an equal number of highest bidders. Consider the following example involving a Dutch auction of $20 billion in 91-day U.S. T-Bills, where the bids (based on yields to maturity) by financial institutions are given in Table 1.1.

Obviously, the Treasury wants to sell as many bills as possible at the highest possible prices. This results in the lowest yields. The bid-to-cover ratio in this illustration is $31.5 billion/$20 billion = 1.525. This means that $11.5 billion in bids will not be successful. Bids will be satisfied from the lowest yield (highest price) until $20 billion in bills have been allocated. The stop-out price will be at a yield of 3.35% and all winners (Citigroup, Wells Fargo, UBS, Deutsche Bank, and JP Morgan Chase) will pay this same

[4]Auctions have a long history. For example, the Roman Empire was sold at auction in C.E. 193 by the Praetorian Guards to Didius Julianus, who reigned as emperor for over two months before being overthrown and executed by Septimius Severus. Earlier references to auctions date back to 500 B.C.E. when the Greek historian Herodotus reported that Babylonians auctioned women as wives.

[5]See O'Callahan (1993) for a discussion of this market.

TABLE 1.1 Treasury Bids Illustration

Citigroup	$2.0 billion at 3.15%
Wells Fargo	$4.5 billion at 3.20%
UBS	$5.5 billion at 3.25%
Deutsche Bank	$7.5 billion at 3.30%
JP Morgan Chase	$5.5 billion at 3.35%
Bank of America	$6.5 billion at 3.40%

price. Actually, JP Morgan Chase will be allocated only $0.5 billion, because its bid filled the $20 billion total being offered.

In the Yankee variation of the Dutch auction, all successful bidders would pay the prices that they bid. While the Yankee variation has the potential to deliver a higher total sales volume, it might discourage bidders from bidding aggressively. See the "winner's curse" below.

Another variant of the Dutch auction was used by Filene's, a department store that was based in Boston. In the store's basement were marked-down goods, each with a price and date attached. The price paid at the register was the price on the tag minus a discount that depended on how much time elapsed since the item was tagged. The longer the item remained unsold on the shelf, the more its price was reduced.

First-price sealed-bid auctions have all bidders simultaneously submit sealed bids so that no bidder knows any of the other bids. The first-price sealed-bid auction does not allow for price discovery until the auction concludes. The winner submits the highest bid and pays the bid price. The winner of a bidding contest in these three auction structures faces the "*winner's curse*" problem if the auctioned item's value is not known with certainty. All bidders' bids are subject to value estimation error. The winner bids the most, and is the most likely to have bid too much for the auctioned object.

The *second-price sealed-bid auction* (Vickrey auction) is identical to the first-price sealed-bid auction, except that the winner pays the highest losing bid rather than his own winning bid. The motivation for the Vickrey auction is to encourage higher bids, in part, by reducing the winner's curse. In a second-price auction, bidders bid more aggressively because a bid raises the probability of winning without increasing the expected cost, which is determined by someone else's bid. A variation of the Vickery auction is used in the Federal Reserve Bank Agency Mortgage-Backed Securities (MBS) markets (we will discuss agency instruments in the next chapter).

Double auctions or *bilateral auctions* are used for the trading of most publicly traded securities in secondary markets (aftermarkets). In double auctions, buyers submit *bids* (maximum purchase prices) and sellers submit offers (minimum selling prices; also called asks) that are ranked from best to worst. These bids and offers create demand and supply profiles for the market. Transactions occur when the highest bid and lowest offer match. A *continuous double auction*, such as that on the New York Stock Exchange,

allows for many transactions on an ongoing basis as the security is continuously being auctioned.

Auction Outcomes

The type of auction structure selected in a market can have a significant impact on submitted bids, winners, final prices, revenues realized by sellers, and information realized in the auction process. However, the *revenue equivalence theorem*, perhaps the most significant result from the game theory of auctions, states just the opposite. The revenue equivalence theorem holds that, under specific restrictions, the auction type (from the listing above) will not affect the auction outcomes. Consider, for example, an environment with perfect and complete information, where all risk-neutral bidders know the values that all bidders place on the object to be auctioned. In this environment, the winner of a first-price sealed-bid auction with infinitesimal bidding increments is the participant who most highly values the auctioned object. The winner will pay a price equal to the highest price submitted by the auction losers. Note that this auction outcome is identical to the second-price sealed-bid auction. In addition, the English and Dutch auctions produce the same outcome in the perfect and complete information scenarios. The key to reconciling this conflict is that the revenue equivalence theorem is based on assumptions of perfect and complete information, with risk-neutral bidders and infinitesimal bid increments. Thus, the analysis of auction type should focus on the nature of violations of these assumptions. Actually, perfect and complete information is not a necessary assumption; it is sufficient that each bidder's value be independent and randomly drawn from a common distribution.

While the revenue equivalence theorem states that under restrictive circumstances, the auction structure will not affect revenue, experimental, statistical, and other evidence suggests otherwise. For example, evidence suggests that there is more overbidding in second-price auctions than in standard English auctions (e.g., see Kagel, Harstad, & Levin, 1987). Thus, one needs to pay close attention to how deviation from revenue equivalence theorem assumptions affects auction outcomes.

Common Value Auctions

A *common value auction*[6] occurs when all bidders place the same value on the item to be auctioned, and that value is known with certainty. Consider the following auction example with three bidders such that each has the opportunity to bid on some random amount of cash between zero and \$1. Suppose that every monetary value between 0 and \$1 is equally likely, such that $E[V] = \$0.50$. Since each bidder has equal access to information, we will refer to this structure as a *symmetric information structure* problem. Without any additional information, risk-neutral bidders might value the random sum of cash as highly as \$0.50; risk-averse bidders will value the random sum less than \$0.50. Thus, we see here that risk aversion will affect valuation of an object of unknown value. Therefore, risk aversion will affect bids and the revenue equivalence theorem will no longer apply. In addition, the

[6]See Kagel and Levin (1986) and Thompson and Wright (2005) for discussion of this problem.

revenue equivalence theorem no longer applies because bids will not only be a function of expected value, they will also depend on information revealed in the bidding process.

Now, suppose that each of the three bidders will obtain a noisy signal, s_1, s_2, and $s_3 \in (0, 1)$, concerning the value of the bundle of cash such that the mean of the signal amounts equal the value of the bundle: $(s_1 + s_2 + s_3)/3 = V$. Now, suppose that Bidder 1's signal is $s_1 = 0.80$. Then, his estimate of the value of the bundle is $E[V_1 \mid s_1 = 0.80]$. If Bidder 1 is risk neutral, \$0.80 is the value that he attributes to the bundle. But, does this mean that Bidder 1 will bid \$0.80 for the bundle?

First, recall the *winner's curse* problem raised earlier. If Bidder 1 wins the auction by bidding \$0.80, this will mean that the other bidders received lower value signals than Bidder 1, indicating that \$0.80, the signal received by Bidder 1, certainly exceeded the bundle value. Thus, winning the auction is a negative signal (ex-post) as to the value of the bundle. If Bidder 1 wins the auction by bidding \$0.80, he will have overbid and will suffer from the winner's curse. This means that, from the perspective of Bidder 1 if he wins, the distribution of bundle value must range from 0 to 0.80 rather than from 0 to 1.00. This is because if Bidder 1 makes the highest bid, it is because his signal was higher than those of his competitors. Thus, the high end of his valuation range will be \$0.80. This means that the anticipated mean signal values received by other bidders should be \$0.40, since their valuation ranges will be from 0 to \$0.80, given that Bidder 1 will submit the highest bid. Thus, based on this information, Bidder 1 should revise his bid for the bundle to $[B_1 \mid s_1 = 0.80] = (\$0.80 + \$0.40 + \$0.40)/3 = \$0.5333$.

All four auction designs will place the bundle with the same winner. However, let us consider Bidder 1 bidding strategies under different auction structures. In a second-price sealed-bid auction, \$0.5333 is an appropriate bid because there is no information revealed in the auction process. In a first-price sealed-bid auction, Bidder 1 should bid less than \$0.5333. This is also true in the Dutch auction. However, additional information is revealed in the bidding process in the English auction, and this information serves to reduce the winner's curse. This information should be used in forming the bidding strategy. Generally speaking, with more than two bidders, the highest bid or revenue realized in the English auction will be greater than or equal to that in the second-price sealed-bid auction, which will be greater than or equal to that in the Dutch auction, which will equal that obtained in the first-price sealed-bid auction. However, English auctions are often more vulnerable to manipulation by shills. Klemperer (1999, 2002) provide a good review of results in auction theory.

Google's Dutch Auction Format

Google, in its widely publicized 2004 IPO, structured a Dutch auction process intending to sell 25.8 million shares of its stock, suggesting bids in the range of \$108 to \$135 per share. This Dutch auction searched bids for a clearing price that enabled it to finally sell 19.6 million shares at the IPO price of \$85, thereby raising \$1.67 billion. The first trade price was \$100.01, rising to over \$300 within a year and over \$500 within three years. Lead underwriters, Morgan Stanley and CS First Boston, collected a 3% commission on the offering rather than the standard 7% fee. Some observers opined that, if successful,

the Google IPO could lead to a transfer of power and fees away from underwriters in favor if issuing firms. However, it is not clear just how successful the IPO was. The IPO price was not as high as anticipated or nearly as high as subsequent trading prices. A later follow-on offering was priced at $295 per share, raising $4.18 billion. On the other hand, the IPO created substantial favorable publicity for the firm and raised fortunes for its owners.

1.5 INTRODUCTION TO MARKET MICROSTRUCTURE

Investors participate in the markets for information, the markets for securities, and the markets for transaction services (Stoll, 2003). *Market microstructure* is concerned with the markets for transaction services. Market microstructure is the area of financial economics that concerns trading and market structure, market rules and fairness, success and failure, and how the design of the market affects the exchange of assets, price formation, and price discovery. Market microstructure is concerned with costs of providing transaction services along with the impact of transactions costs on the short-run behavior of securities prices (see Stoll, 2003). The market structure is the physical (or virtual) composition of the market along with its information systems and trading rules. Market microstructure examines *latency*, the amount of time that elapses between when a quote or an order is placed by a trader and when that order is actually visible to the market. Microstructure also concerns transactions costs and the impact of transactions costs on the behavior of securities prices. Transactions costs are often reflected in bid-ask spreads, security price mark-ups, and in commissions paid to brokers. Generally, the best market is that which has the lowest transactions costs, facilitates the fastest trades, results in the fairest prices, disseminates price information most efficiently, and provides for the greatest liquidity. A market is said to be liquid when prospective purchasers and sellers can transact on a timely basis with little cost or adverse price impact.

The basic components of a securities market are investors, brokers and dealers who make markets and facilitate trading, and the market facility, physical and virtual, within which trading takes place. Investors include individuals and institutions. Some market participants are price takers (typically individual investors and institutions that invest for the longer term) and others are price seekers (day traders and market makers). Brokers act as agents, processing trades on behalf of client accounts on a commission basis. Upstairs brokers deal directly with investors, and downstairs brokers execute transactions on trading floors. Dealers trade on their own accounts to secure trading profits.

Market Execution Structures

Securities markets are categorized by their *execution systems*, that is, their procedures for matching buyers to sellers. The primary execution systems are *quote-driven markets* where dealers post quotes and participate on at least one side of every trade, *order-driven markets* where traders can trade without the intermediation of dealers, *brokered markets* where many blocks (10,000 or more shares, as defined by the NYSE or 25% of daily volume

elsewhere) are broker negotiated, and *hybrid markets*. Most over-the-counter (OTC) currency and bond markets are primarily quote driven, and most stock exchange, futures, and options exchanges are primarily order driven. Quote-driven markets display quotations of specialists and market makers while order-driven markets display all quotes. In a pure quote market, dealer spreads are the source of market maker profits, whereas in a pure quote market, there are no dealers.

Actually, to some extent, most markets are hybrids. For example, the NYSE uses designated market makers (formerly specialists) who can act as dealers on transactions and brokers may execute block transactions on exchanges. Stock markets around the world, with the most prominent exception being the OTC markets in the United States, are more likely to be order driven. This is because such markets require less human intervention, and tend to be cheaper to run. However, the lack of a dealer book, improved transparency, and dealer quotes tend to reduce liquidity in such markets, and may drive orders away from them to quote-driven markets.

Continuous and Call Markets

Should markets facilitate trading anytime they are open or is it more important to allow orders to accumulate for more competitive order matching? Continuous markets allow for orders to be executed whenever the market is open, whereas call markets allow for trading of securities only at predetermined times or intervals during the trading day. Continuous markets allow for an order execution whenever a bid and offer match or *cross* (higher bid than offer) one another during the trading day. During the course of a trading day, the NYSE and Nasdaq typically act as continuous markets as order matching engines match buy and sell orders against one another. Continuous markets provide for *immediacy*, enabling orders to be executed quickly whenever markets are open.

However, in a call market, orders for a given security are *batched* for simultaneous execution when the market is called, and for those orders to be matched by a certain set of rules or algorithm for execution. Orders to buy and orders to sell a particular security accumulate in the market until the start of the trading session, at which time, orders are matched and securities are traded. The market maker or specialist, if one or more exist in the market, seeks to find a market clearing price by matching buy orders against sell orders so as to minimize the order imbalance. The NYSE and Nasdaq function more like a call market during daily market openings in which purchase and sell orders are batched and matched for market clearing, facilitating price discovery after a period of nontrading. Crossing networks, which can be particularly appealing to investors placing infrequent large orders, also function as call markets, generally with prices calculated from those in other markets.

Both continuous and call markets have long histories and clear relative advantages. Advantages of a continuous market include:

- Continuous markets provide liquidity for investors who wish to trade outside of the called trading sessions.
- Continuous markets allow for frequent trading that can allow for continuous price discovery.

- Continuous markets allow for simultaneous trades in different trading venues.
- Continuous markets post bid and offer quotes on a continuous basis, improving price certainty for investors seeking to trade.

Advantages of a call market include:

- Liquidity is enhanced in call markets during trading sessions due to the accumulation of orders for a security along with the investor focus on that given security at the time of the auction.
- Call markets can improve liquidity for less active markets such as those for less liquid stocks or during less active periods such as after the closing of normal trading activity.
- Large orders experience less *market impact* (price increases due to excess demand or price decreases due to excess supply) in call markets due to the accumulation and potential balancing of orders prior to the market session.
- Call markets do not require that trading apparatus be available to all securities during the course of the trading day as do continuous markets.
- Call markets diminish the speed advantages of participants with speed or other technological advantages such as high frequency traders.
- If trading sessions are relatively frequent, price discovery can be more effective in a call market as only a single equilibrium price results from the auction.
- In a call market, prices need not be affected by the sequence in which orders arrive.

1.6 ORDERS, LIQUIDITY AND DEPTH

Orders are specific trade instructions placed with brokers by traders without direct access to trading arenas. The typical brokerage will accept and place a number of types of orders for clients. In fact, the various exchanges and alternative trading systems in the U.S. will accept over a hundred different types of orders, not counting permutations of the basic types. These order types include some specifically designed for high frequency traders (HFT), others specifically designed to protect against HFT strategies, orders designed to be executed at market opens and closes, and others intended to control which market will be used for execution. Among the more common types of orders are the following:

- *Market order*: Here, the broker is instructed to execute the order at the best price available in the market.
- *Limit order*: An upper price limit is placed for a buy order; the broker will not buy at a price above this limit. A lower price limit is placed for a sell order.
- *Stop order*: Here, the broker is instructed to place the buy order once the price has risen above a given level; in the case of the stop-sell (or stop-loss) order, the broker sells once the price of the security has fallen beneath a given level. Stop-loss orders are often intended to protect against stock price declines.
- *Day order*: If not executed by the end of the day, this order is canceled.
- *Good till canceled order*: This order is good until canceled.
- *Not held order:* Here, the broker is not obliged to execute while he is attempting to obtain a better price for his client.

- *Fill or kill order:* This order must be filled in its entirety immediately, or it is canceled.
- *Immediate or cancel order:* This order is immediately executed to the extent possible; unexecuted amounts are canceled.
- *Market-on-close (MOC) order:* This order must execute at the market's end-of-day closing auction, irrespective of price.
- *Limit-on-close (LOC) order:* This order must execute at the daily closing auction only if the closing price is at least as good as the limit price set by the client.

This list of order types is far from complete and some brokers and exchanges will not accept all of these types of orders. Order types are not mutually exclusive; for example, a good-till-canceled, limit buy order is a legitimate order type as is a stop limit sell order. The stop limit sell order authorizes the broker to initiate the sell order once its price drops to the stop trigger, but only if the limit price can be realized for the sale.

Liquidity

Liquidity refers to an asset's ability to be easily purchased or sold without causing significant change in the price of the asset. Liquid assets can be traded quickly, with low transactions costs, at any time and with little impact on the asset's price. Markets with large numbers of active participants and few constraints on trade are more likely to have greater liquidity. Bid–offer spreads are generally considered to be quick indicators of liquidity, with narrow spreads suggesting that price impacts of trading will not be severe. Black (1971) described liquidity as follows:

1. There are always bid and asked prices for the investor who wants to buy or sell small amounts of stock immediately.
2. The difference between the bid and asked prices (the spread) is always small.
3. An investor who is buying or selling a large amount of stock, in the absence of special information, can expect to do so over a long period of time at a price not very different, on average, from the current market price.
4. An investor can buy or sell a large block of stock immediately, but at a premium or discount that depends on the size of the block.
5. The larger the block, the larger the premium or discount. In other words, a liquid market is a continuous market, in the sense that almost any amount of stock can be bought or sold immediately, and an efficient market, in the sense that small amounts of stock can always be bought and sold very near the current market price, and in the sense that large amounts can be bought or sold over long periods of time at prices that, on average, are very near the current market price.

Kyle (1985) characterized three dimensions of liquidity. The first is *width* (also known as tightness), which is simply the bid–offer spread. The second is *market depth*, which refers to a market's ability to process and execute a large order without substantially impacting its price. The third dimension of liquidity is *slippage* (also known as market impact, price impact, or market resilience), which indicates the speed with which the price pressure resulting from a noninformative trade execution is dissipated (the price returns to normal).

TABLE 1.2 Market Depth

Market A		Market B	
# Shares	Offer	# Shares	Offer
1000	50.00	2000	50.00
2000	50.03	1000	50.01
1000	50.05	2000	50.03
2000	50.06	2000	50.04
3000	50.07	2000	50.05
1000	50.09	3000	50.05

Normally, markets with larger numbers of active participants have more depth than thin markets. Suppose, for example, that there are two competing markets for Stock X with the following offer prices (central limit order book) for Stock X as depicted in Table 1.2. Suppose that the last transaction for Company X stock was at a price of $50.00. Further suppose that an investor places a market order to buy 5000 shares of Company X stock. In Market A, the investor will obtain 1000 shares for $50.00, 2000 for $50.03, 1000 for $50.05, and 1000 for $50.06. The final price rises to $50.06. In Market B, the investor will obtain 2000 shares for $50.00, 1000 for $50.01, and 2000 for $50.03. The final price in Market B rises to $50.03, less than Market A. Thus, Market B has more depth than Market A, at least with respect to the stated demand for the stock. We will focus more on liquidity in Chapters 4 and 6.

Order Precedence Rules

The typical double auction will have many bids and ask prices for a given security at any point in time. Whose quotations will be the next to be executed or will have priority in the next transaction? *Order precedence rules* concern which traders can place bids and offer quotations with greatest priority for execution and which can accept the quotations of other traders. In almost all cases, the primary precedence is *price priority*. The participant with the highest bid or lowest ask has priority on an execution. However, there will be many instances where multiple traders have made quotations at the same price. Which quotations receive secondary priority? Some markets give priority to the first trader to make that particular quotation if that quote improved the previous best quote. This is called *time priority*. The trader who last improved the price is normally the first to post a quote at the current best price. Time priority is intended to encourage traders to aggressively improve their quotations and is most effective when the tick size (minimum increment for quote differences) is larger. However, some markets will provide for public order precedence, meaning that orders placed by the general public will have priority over otherwise competing orders placed by exchange members. These public precedence rules are intended to enhance the credibility and perceived fairness of markets in the eyes of the general public.

Even price priority can be violated when a security trades in multiple markets. For example, consider the scenario where the best bids on a stock are $49.95 for 100 shares and $49.94 in another market for 1000. A prospective seller wishes to sell 1000 shares. In some instances, under "best execution" rules (discussed in Chapters 4 and 5), rather than break up the 1000 sell order, the broker will *trade through* the $49.95 order in one market, executing the entire transaction for 1000 shares at $49.94 in the second market.[7] In most instances, this cannot be accomplished legally in a single exchange market. Sometimes time or even price priorities are violated when a broker routes an order to a market in exchange for *payment for order flow*. In this scenario, the broker might receive a payment from a satellite market or market maker such as Citadel Securities for routing the transaction to it rather than to the principal market for the security. We will discuss payment for order flow in more detail later in this chapter and in Chapters 2–5 and 13.

Unlike most other exchanges, the NYSE employs a model of precedence known as *parity and priority*, in which price priority is still the primary rule when executing multiple orders. However, with multiple quotes at the same price, time priority no longer takes secondary precedence for execution. Instead, orders at the same price are allocated against a sufficiently large counterparty trade with the original setter of the best quote being first allocated 15% (rounded to the nearest round lot) of the execution, then all orders except those in the Electronic Book (off-floor orders from the public), if shares remain to be traded, including those from the original price setter (even if in the Electronic Book), the DMM, and floor brokers participating on a parity (equal) basis for their displayed orders. After these member orders are executed, orders remaining in the Electronic Book are executed once the original price setter order is fulfilled or cancelled. Small and discretionary orders do complicate the priority model somewhat.

Whereas time priority rewards faster traders with execution priority, the NYSE claims to level the playing field by allowing for more sharing of executions. This may serve to improve execution fill rates and reduce transactions costs for retail traders, encouraging broader market participation. These claims might be somewhat controversial (see, for example, Battalio, Jennings, and McDonald [2021]), and the NYSE priority rules might increase trading costs to the general public, partly because of the priority advantage enjoyed by floor brokers over the Electronic Book (which includes retail investors) and partly because most retail trades route to markets other than the NYSE.

1.7 DAY TRADING

The Internal Revenue Service (IRS) defines day traders to be individuals who have all of the following three characteristics: (1) The trader maintains substantial trading activity. The trader buys and sells frequently (10 to 20 daily trades should be sufficient) and trading is a primary source of income for the trader. (2) The trader's trading activity is sustained on a regular and continued (one year minimum) basis. (3) The trader seeks to profit

[7] A trade through occurs when a broker executes a trade in a given market even a better price exists in another market. We will discuss trade through scenarios and regulations in Chapters 5 and 13.

from short-term stock price fluctuations rather than dividends, capital appreciation, and interest.

The day trader should complete a year in advance IRS Form 3115 Application for a Change of Accounting Method.[8] This is crucial to avoid difficulties with the wash sale rule. The purpose for completing this form is to seek permission to use the mark-to-market accounting technique. In addition, this designation will strengthen the trader's argument that she is a for-profit day trader, allowing for better expense deduction possibilities, such as departing from the 2% miscellaneous threshold and the at-risk rules. In addition, at-home day traders should consider whether they qualify for home office and other depreciation deductions.

FINRA (the Financial Industry Regulatory Authority; formerly NASD, the National Association of Securities Dealers), based on its Rule 2520, defines the day trader somewhat differently. Anyone who executes buy and sell transactions on the same margin account on the same day is said to be day trading. A *pattern day trader* executes four or more of these round-trip transactions within five consecutive business days.[9] Pattern day traders are required to maintain only 25% margin requirements rather than the 50% maintained by other noninstitutional traders. Rule 2520 requires maintenance of $25,000 of equity in the margin account to take advantage of this exception, but the pattern day trader can margin this account by four times, rather than the usual two times based on the 50% rule for other individual investors.

For other individual investors, Federal Reserve Board (FED) Regulation T requires 50% initial margin along with 25% maintenance margin. Thus, such an investor would have to put up $5000 in cash to purchase $10,000 in shares. This would give the investor $5000 in equity position in the shares, or a 50% initial margin. However, if the value of the shares were to drop to $6000, the investor's equity would drop to $1000, or to 16.67% of the current investment (1000/6000), well below the 25% maintenance margin. The broker would then need to require the investor to post an additional $500 ($25\% \times \$6000 - \$1000$) to fulfill the margin requirement.

Brokerage clients should carefully read their *hypothecation agreements* that define terms of their margin accounts, including unauthorized liquidations of shares purchased on margin and loaning of shares to other investors who may wish to short sell them. The hypothecation agreement will detail terms concerning pledges of client assets (how they might be used as collateral) as well as how the broker might pledge the client's assets for its own obligations. Equally important, they should be very familiar with the requirements imposed by FED Regulation T and FINRA Rule 2520.

While the Securities and Exchange Commission (SEC) and the IRS have fairly specific definitions for day traders, other definitions exist as well. For example, some people distinguish between day traders and swing traders. Some say that swing traders usually hold a stock or a position from one day to a week and tend to concentrate on just a few selected

[8]File Form 3115 with an income tax return prior to requesting the change of accounting method. These are not "stand-alone" forms. Consider making such a filing after forming a corporation (such as an LLC) for trading.

[9]Unless these trades summed to less than 6% or fewer of all trades made over these five days.

stocks or positions. In this context, day traders neutralize their positions at the end of each trading day.

Online Brokers and Direct Access Trading

Selection of a broker, or brokers, is a crucial decision for the day trader. *Barron's* conducts and publishes results of surveys rating the various online brokers. When making this decision, the trader should consider a number of factors, including commissions and other fees, trade execution quality, account balance minimums, margin and other interest rates, customer service and live-agent telephone access, bank services, information, research and analytical services, trading platform, level of real-time quotes, and investment selection (some brokers won't work with mutual funds, IPOs, or bonds). Other considerations might include whether the broker is or has been targeted by the SEC, FINRA, or NASD for investigation for significant and persistent violations, along with the outcomes of these investigations. How the broker routes orders and the extent to which it accepts payment for order flow might also be important considerations.

Full-service brokers such as Oppenheimer, Raymond James, and UBS provide a wide array of services to their clients, including trade execution, advice, market research, and so on. They typically are compensated with commissions on securities transactions, interest on uninvested cash balances and might impose periodic account maintenance charges as well. *Discount brokers* such as Interactive Brokers, TD Ameritrade, E*Trade, TradeKing, and ScotTrade provide for online trade execution, and may or may not provide other services as well. For most long-term investors and some day traders, services provided by one or more of these types of brokerage firms are perfectly appropriate. Most of the firms listed here provide more than a single type or level of service to a wide array of different types of clients.

Trade execution commissions are an important source of revenues for many brokers. However, this source of revenues has taken on less importance in recent years as brokers have relied more on producing revenues from:

- interest from lending clients' idle cash balances;
- margin interest—that is, charging clients interest to purchase securities on margin;
- fees from loaning clients' securities to short sell;
- sales of their own mutual funds and ETFs;
- payment for order flow, a huge source of revenue for Robinhood Financial and Ameritrade;
- subscription fees for premium services such as research reports and NASDAQ Level II market data (e.g., Robinhood Gold);
- portfolio management and investment advisory fees;
- selling educational, advisory, and other financial services; and
- various administrative and service fees such as for wire transfers, software, yearly administrative fees, signature guarantees, back-end fees for account transfers, and account liquidations.

Massa (2019) reported that as of August 2019, Charles Schwab clients' accounts totaled about $3.7 trillion, with about $265 billion of this sum earning interest for Schwab rather

than for clients, accounting for as much as 45% of its revenues. Some brokerage firms have used "zero commission" promotions to attract new customers, which then might pay higher stock-trading commissions after the promotion periods end. Retail customer orders can be especially desirable for many brokers and the dealers from which they receive payments for order flow because retail clients are assumed to be uninformed, creating no adverse selection risk for the broker's dealers or market makers. Accordingly, many retail brokers route nearly all their customer market orders to market makers and internalizers seeking payment for order flow arrangements. Brokerage firms might use low commissions to "hook" clients with the hope that they can sell them premium services such as wealth management services.

In many instances, high-volume traders will require faster and better trade execution. *Direct access trading systems* may provide for faster and potentially superior execution for such traders. Direct access trading through firms such as such as Interactive Brokers, Questrade, and TD America's Thinkorswim enables traders to execute transactions directly with market makers and designated market makers on the NYSE, NASDAQ, and certain ECNs, eliminating the broker from direct transaction participation. The trader may have more control over routing the transaction, avoiding issues related to payment for order flow and *slippage* (movement in the security price against the trader caused by the trader's impact on supply or demand). While transactions through brokers can execute within a fraction of a second, transactions in many markets take at least seconds and some take minutes or longer. Most direct access transactions on active exchanges execute within fractions of a second. Direct-access trading fees are typically volume based, include exchange and other market fees, and may include fixed platform and software fees. However, total fees can actually be higher than those charged by the deepest discount brokers. In addition, more knowledge is likely to be required on the part of the trader, and high trading volume is likely to be necessary to make this method of trading more cost effective.

Trading Platforms

A trading platform is a computer system used to place or route quotes and transactions through a network to a financial intermediary or market. Trading platforms can be either software-based (e.g., TradeStation, E*Trade Pro, Bloomberg EMSX, and Litespeed Trader), or web-based such as those provided by most brokers (e.g., Charles Schwab Active Trader, Thinkorswim, and E*Trade). Software-based platforms are usually integrated with analytical tools, although many web-based platforms are as well. TradeStation provides trading technology for stocks, options, futures, and forex (foreign exchange, or currencies), along with tools for creating strategies and testing, notably, including a simulation for back testing. Trading platforms can monitor markets and can often be programmed to automatically execute trading strategies based on the customer's custom trading rules. This can mean that the trading platform can be customized for the customer's own trading algorithms.

While trading platforms are easily obtained from brokers or developers of relevant software, there are a number of advantages to the trader who creates her own platform. In fact, many brokers and trading arenas are equipped to feed data into and accept quotes and execute transactions through Excel spreadsheets. Thus, a trader can create a spreadsheet designed to receive market data such as quotes and recent transactions and program

in his own trading rules or algorithms to transmit quotes and execute transactions. Such platforms can quickly analyze market data, respond to trading rules, and provide for simultaneous transmission of multiple quotes and executions. Speed, accuracy, and efficiency can be enhanced considerably with the preparation of appropriate spreadsheet macros. For example, macros can be created that compare bid and offer quotes for a number of different securities, complete computations, and then transmit quotes or execute transactions based on "if/then" statements reflecting trading rules. These macros can include buttons to follow rules and can scan data and execute transactions in the trader's absence. In addition, a custom spreadsheet-based trading platform is obviously very flexible and, often with little adjustment, can continue to be used when the trader switches brokers or trades new instruments. There are a number of firms that can provide assistance with developing spreadsheet-based trading platforms, although many traders should be sufficiently competent to develop platforms without significant assistance.

Quotations and Price Data

The day trader will need timely access price quotes. Markets where securities are traded usually retain ownership of quotation, price, and transaction data and sell these data to interested customers. In fact, in recent years, the single largest source of revenue to major exchanges such as the NYSE has been from the sale of price, volume, and quote data. Real-time quotes are available to traders as quickly as they can be transmitted and displayed; otherwise they are said to be delayed. Delayed quotations are usually less expensive than real-time quotations data. However, in a trading environment where milliseconds (thousandths of a second) or even microseconds (millionths of a second) matter, what exactly is real-time data? In theory, real-time data display exactly as quotes are placed and transactions are executed. However, quotes and transactions data cannot be made available to all traders instantaneously. Different data and vendor services provide these data using different technologies from different locations. Traders compete to obtain these data as quickly as possible and vendors compete to provide it as quickly as possible. Microsecond and even millisecond delays are to be expected, and can easily spoil many trading strategies.

More extensive real-time quotations data are more expensive than less extensive data. Level I quote access displays the best bid and offer prices (inside quote or BBO: best-bid-offer) and, in some cases, quote sizes. Level II quotes display the same along with other quotes in descending order for the best bids and ascending for the best offers along with market symbols for each (see Figure 1.1). Symbols include NBBO for the national best bid offer (best buy and sell quotations), markets, recent transactions, and so on. Market participant IDs are omitted. Level II access provides the order book (or TotalView access for all quotes) and is necessary for most trading strategies. Almost all brokers provide their clients with free Level I real-time quotes, and most now provide free Level II quotes as well. Some may also provide free Level III quotes, but this varies over time. IEX and several other exchanges (see Chapters 2 and 4) offer real-time data including free depth-of-book details to many of their clients and subscribers. Level III quotes, offered to NYSE, OTCBB (Over the Counter Bulletin Board, described in Chapter 2) and NASDAQ members, enable

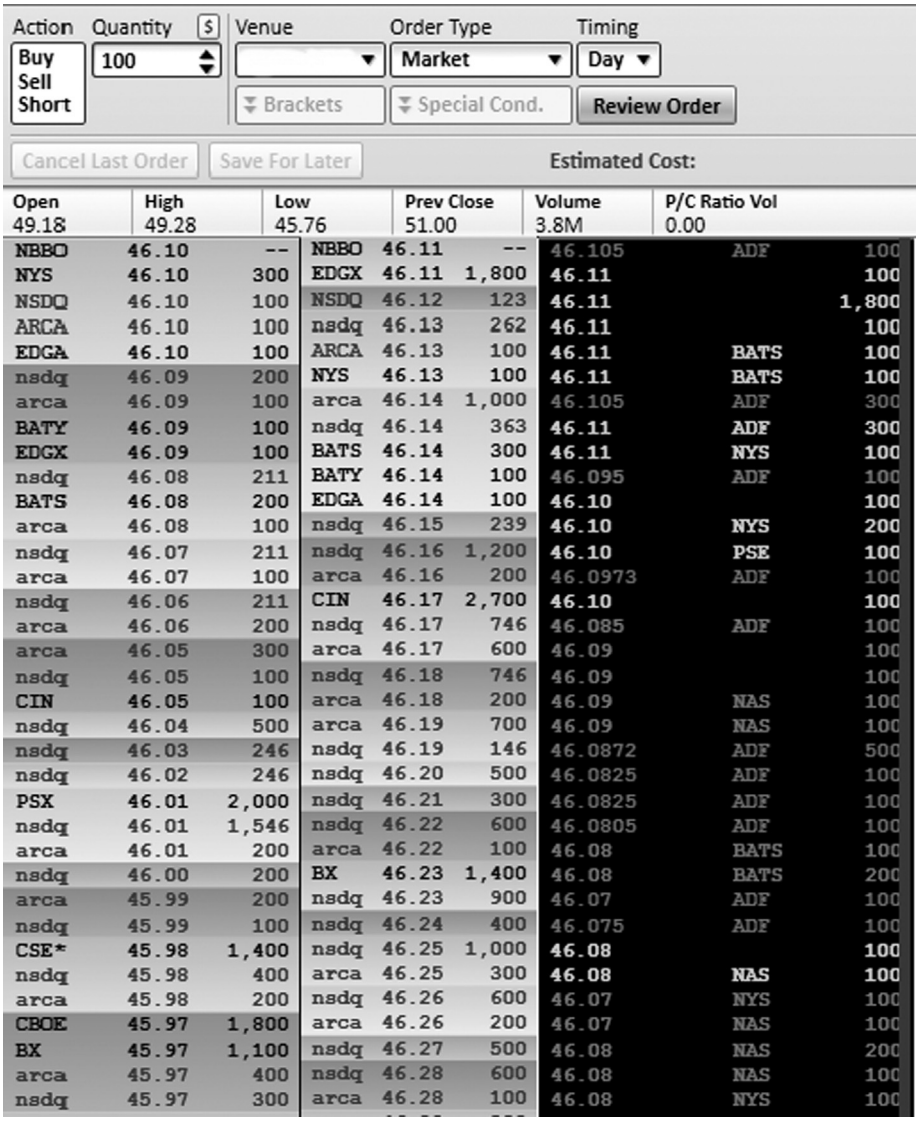


FIGURE 1.1 BATS Level II Quotes, MSFT (Microsoft), August 17, 2011.

traders to have direct access to enter and revise quotes, execute orders and receive trade confirmations. NASDAQ's SuperMontage TotalView provides more detail on the depth of data than Level II, enabling traders to view market makers' quotes that are not as good as their best. Figure 1.2 provides a sample screen from SuperMontage TotalView. Note that multiple quotations are frequently provided by several of the MPIDs (market participant IDs).

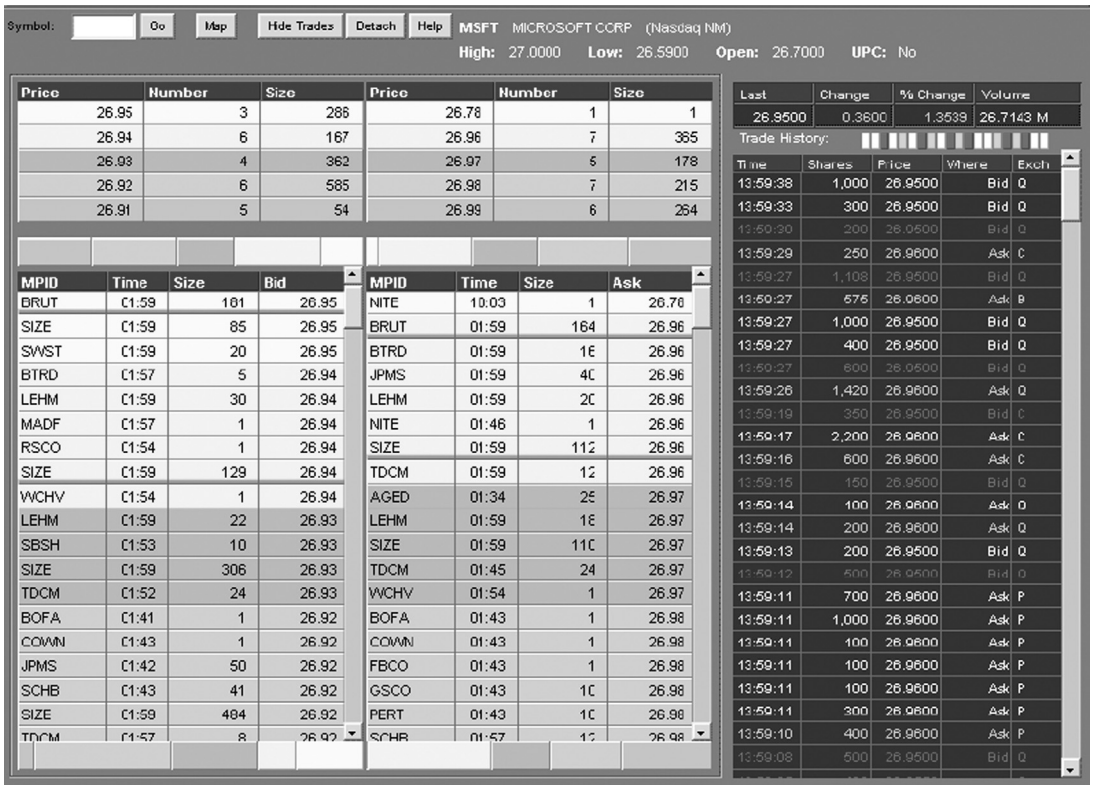


FIGURE 1.2 Screen from NASDAQ's SuperMontage TotalView.

One of the most popular sources of market information and price data is Bloomberg, which offers real-time data and news through *Bloomberg Professional* and access to these data through its Bloomberg terminals. Costs vary, but as of 2018, single machine access can be licensed for roughly $24,000 per year, and interestingly, with only relatively insignificant volume discounts. The system provides data, news access, analytical tools, email, and trade processing systems, and is used by over a quarter million professionals worldwide. There are many, and often less expensive, competitors, including Refinitiv Eikon from Thomson Reuters, FactSet Research Systems, and information content vendor Factiva (Dow Jones). Less expensive quotations systems, such as eSignal and MetaStock, offer prices and quotations for as little as $100 per month, and sometimes provide charts and related services.

More specialized data systems are available as well. Real-time data for stock, currency, and futures markets can be obtained via Internet, wireless, or satellite feeds. Specialized providers such as IQFeed and e-Signal offer quotes and recent executions to traders and can be linked to spreadsheet-based trading platforms. On a fee basis, information providers such as Dow Jones and Thomson Reuters can offer electronically "tagged" news products that can be picked up by computer algorithms to trigger programmed trading decisions.

Trading Arcades and Prop Shops

A *trading arcade*, sometimes, and perhaps incorrectly referred to as a proprietary trading firm or prop shop, is a physical (and sometimes virtual) location for traders to work and trade. The trader in the arcade might lease office space, a desk, a trading platform, computers and screens, analytical services, access to market data and news services such as Bloomberg or Reuters, order routing technology, clearing and settlement services and office facilities, in an effort to benefit from reduced expenses and economies of scale. Increasingly, no physical space is provided; instead, traders can work from home. Some trading arcades will provide some or all of the capital to be traded, with or without interest. The trader can usually expect to receive reduced brokerage commissions and other trading costs. In addition, the arcade might charge for training services and receive payments from exchanges for order flow. The typical trading arcade will promote itself as "one-stop" shopping for independent traders.

The distinction between trading arcades and prop shops is often blurred, though trading arcades tend to focus more on providing trading facilities to independent traders while prop shops tend to focus on building long-term employee or partnership arrangements, providing internships, free or paid training programs and close ongoing support to traders. Many prop shops will focus on a variety of market-making activities, and function similar to traditional proprietary trading firms and trading departments in major financial institutions. Some prop shops and trading arcades will provide capital to traders in exchange for a split in trading profits. Firms that provide capital and receive all or most of trading profits are more likely to be referred to as prop shops while true arcades simply provide trading facilities to traders who trade capital in their own accounts.

Many trading arcades were created by floor traders who were rendered obsolete when their trading environments transformed into electronic markets. These traders transitioned from physical trading floors where physical stamina and ability were advantages to electronic arenas where keystroking ability such as those gain from playing video games provided an advantage. Some arcades have focused on working with amateur traders who have gone through periods of unemployment. In some cases, they have provided training to prospective traders, often for a fee, after which many traders failed in their trading efforts. Trading arcades are frequently short-lived businesses, although some of the longer-lived arcades have included those operated by Maverick Trading. Prop shops are more likely to invest heavily into their affiliates and employees for traders. Well-established prop shops and larger financial institutions with prop shop units include Jane Street Capital, Hudson River Trading, Group One Trading, Susquehanna International Group, Geneva Trading, DRW Trading, and Citadel.

1.8 TRADING AND THE BROADER ECONOMY

The financial services industry plays a crucial role in national and world economies by creating, trading, and settling financial instruments, which are associated with the capital needed for production of goods and services, for shifting funds from "surplus agents" to "deficit agents" and for risk mitigation and shifting. Growth in financial service sectors often accompanies growth in real production sectors, as was clear in the U.S. during the

1920s era, which was characterized by technological improvements and the post 1980s IT growth era. However, as we will discuss shortly, the U.S. economy also experienced significant productive growth during the post-War period 1945–1970 with only a modest share of growth in the financial services sectors. Nevertheless, since issuing and trading of financial instruments is key to planning, production and distribution in a capitalist economy, it is important to evaluate the important role of trading.

The Scale of Financial Markets in the U.S. Economy

Tremendous financial growth of financial services industries has helped fuel significant overall economic growth, produced huge compensation packages for financial executives and even larger profits for financial firms. Philippon (2013) reported that "total compensation of financial intermediaries (profits, wages, salary and bonuses) as a proportion of GDP [adjusted for government defense spending] is at an all-time high. We see in Table 1.3 that

TABLE 1.3 Value Added by Industry as a Percentage of Gross Domestic Product

	1947	1955	1965	1975	1985	1995	2005	2015
Gross domestic product	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Agriculture, forestry, fishing, and hunting	8.0	4.4	3.0	3.0	1.8	1.2	1.0	1.0
Mining (including oil & gas extraction)	2.3	2.4	1.5	2.0	2.5	1.0	1.7	1.8
Utilities	1.4	1.9	2.1	2.3	2.6	2.3	1.5	1.6
Construction	3.6	4.3	4.5	4.4	4.1	3.9	5.0	4.1
Manufacturing	25.4	27.6	25.7	20.9	18.5	16.5	13.0	12.0
Wholesale trade	6.2	5.9	6.3	6.8	6.2	6.0	5.8	6.1
Retail trade	9.3	7.8	7.6	7.6	7.2	6.9	6.5	5.9
Transportation and warehousing	5.7	4.9	4.0	3.6	3.2	3.1	2.9	3.0
Information (incl. publishing, software, motion picture & sound recording, telecom)	3.1	3.3	3.7	3.9	4.6	4.7	4.9	4.7
Finance and insurance	2.3	3.1	3.5	4.3	5.4	6.4	7.6	7.2
Real estate and rental and leasing	7.9	9.3	10.3	10.4	11.7	12.0	12.6	13.1
Professional and business services (including legal, computer, technical, management)	3.3	3.8	4.5	5.2	7.4	9.1	11.0	12.2
Educational services, health care, and social assistance	1.8	2.2	2.9	4.3	5.2	6.9	7.3	8.3
Arts, entertainment, recreation, accommodation, and food services	3.2	2.7	2.7	2.8	3.1	3.4	3.7	3.9
Other services, except government	3.0	2.7	2.7	2.4	2.4	2.6	2.4	2.2
Government (including military, non-military, state & local, government enterprises)	13.5	13.7	15.0	16.2	14.1	14.1	13.2	13.0

Adapted from: Bureau of Economic Analysis; Release Date: November 3, 2016.

this proportion was 7.2% in 2015, based on 2015 GDP of approximately $18 trillion and the share of the financial industry at about $1.3 trillion (the latter two nominal figures are not listed in Table 1.3). This proportion had risen from a 1947 post-war level of 2.3%, based on 1947 GDP of $250 billion and the financial industry's share of $5.9 billion. Thus, this tremendous growth in U.S. economic output has been accompanied by more than triple the growth rate in the financial services industries, although much of this financial sector growth has occurred after 1980.

While the financial sector's share of GDP has set new records since 2000, it was nearly as high during the 1920s before collapsing in the 1930s and 1940s. As we noted above, the growth of the financial services industries was not always correlated with the economic growth rate, such as during the relatively strong post-War economy before the 1980s, when regulatory and other structural factors restrained financial sector growth. With some growth in the 1970s and larger growth rates afterwards, we see from Table 1.3 that the U.S. financial sector has more than tripled in size relative to GDP. Significant growth rates have also been recorded in U.K., Australian, Canadian, and Swiss economies.

The Scale of Trading in Financial Sectors

Trading and liquidity are essential for healthy financial markets, for facilitating capital acquisition by productive institutions and for improving the flow of information. There is no doubt that trading has reached unprecedented levels in recent years. For example, in the U.S., total dollar trade in S&P 500 stocks reached $29.5 trillion, close to 1.7 times U.S. GDP early in the 21st century (see Collin-Dufresne & Daniel, 2014). Table 1.4 suggests that

TABLE 1.4 Stock Market Turnover Ratio (%)

	1975	1980	1985	1990	1995	2000	2005	2010	2020
World	8	16.8	26.3	24.6	18.2	24.8	16.9	15.1	104
Australia		10.1	29.9	33.1	40.5	52.8	75.1	75.9	71.3
Canada	10.7	10.3	14.8	13.4	29.3	56.2	67.6	69.5	38.3
China					109	158	91.9	205	258
France	16.6	16.8	29.3	34.3	42	78.6	76.5	71.6	63.6
Germany	20.4	18.6	67.4	139	101	146	108	113	79.4
Italy				24.6	44.6	141	132	115	341
Japan	43.8	54.2	57.1	43.8	30.7	63.1	108	116	94.3
Switzerland		189	351	222	87.1	90.9	6.42	74.8	65.3
United Kingdom	41.3	104	26.3	34.8	39.9	68.9	59.1	103	66.0
United States	20.1	34.6	75.3	62.3	85.4	198	153	222	108

Compiled from various sources listed in this subsection. 2020 figures were taken from the closest year for which data were available.

equity trading, as evidenced by market turnover in many countries has dramatically increased. The U.S. experienced equity market turnover of 108.5% in 2018, meaning that annual trading in U.S. equities was about 1.08 times the total value of outstanding shares (World Bank, 2021). Table 1.4, compiled from various sources listed in this paragraph, suggests that trading activity in certain other countries, including China, was even higher. Turkey led the world in stock market turnover in 2020 with 365.8% (World Bank, 2021). Even more strikingly, the Bank for International Settlements (2016) found that average worldwide *daily* trading volume in foreign exchange for April 2016 was $5.6 trillion, not substantially different from other months in recent years. In commodities markets, volumes (open interest) of exchange-traded derivative contracts have increased enormously, and are now often 20 to 30 times greater than the actual physical production of underlying commodities, with similar volumes in over-the-counter trade (Silvennoinen & Thorp, 2013). Thus, increased levels of trading were pervasive across financial instruments and countries.

Financialization

Clearly, financial intermediation and trading in particular have played ever-increasing roles in national and world economies. With substantial resources and talents directed towards financial services, we might expect or at least hope that that productive firms have better and less expensive access to capital than ever. So, have these increasingly important roles of the finance sector actually delivered value-added to real economies? Or, has financial sector growth crowded out competing real economic sector growth? To what extent has financial sector growth improved human welfare and heightened economic stability?

Krippner (2005) and Arrighi (1994) defined *financialization* as"a pattern of accumulation in which profits accrue primarily through financial channels rather than through trade and commodity production." We now ask whether our increased economic focus on the financial sector has improved our ability to direct resources to production or actually shifted resources away from real production.

Philippon (2015), in his examination of these issues, characterizes the cost of financial intermediation as "the sum of all spreads and fees paid by non-financial agents to financial intermediaries," which is "also the sum of all profits and wages in the finance industry." Philippon estimated these costs over each of 142 years in the U.S., finding them to range around 1.5%−2% of intermediated assets, showing constant rather than increasing returns to scale and remarkable consistency over time despite drastic improvements in technologies. These results are surprising. Firms were able to obtain needed capital at pretty much the same per-unit cost in 1870 as in 2012, despite huge growth and prodigious applications of new technologies in finance industries. How is it that the per-unit costs to firms seeking financial services do not decrease, even as the technological innovation would seem to reduce the financial institutional costs for providing these services? Philippon argues that these efficiencies were consumed as compensation and profits by finance professionals and financial institutions rather than passed on to firms in real sectors as either service improvements or cost reductions. Thus, he suggests, financialization did not so much improve the process of

intermediating capital between surplus and deficit agents, but instead was associated with increased creation and trading of financial instruments, which served to increase compensation to financial executives and profits to financial institutions.

However, the compensation captured by finance professionals and the profits realized by their employers do not fully capture the costs of financialization to society. Consider, for example, the investment of $300 million into the Spread Networks high-speed fiber optic cable that, starting in 2010, enabled competing financial traders to reduce their communication times between markets in Chicago and New York from 16 to 13 ms (a very fast blink of a human eye is approximately 200 ms). This installation was quickly followed in 2011 by microwave technology (constructed by data networks provider McKay Brothers) that reduced the time to 9 ms, and with further investment, 8.5 and then 8.1 ms (Budish, Camton, & Shim, 2015). Similar communications investments were made connecting Tokyo and Singapore markets, New York and London (the Hibernia Express 4600 km transatlantic submarine cable) in 58.95 ms, and numerous market centers around the globe. It is not easy to believe that such minute improvements in trading speed reduce the cost of capital for real-sector firms.

The technological arms race is probably dwarfed by the race for brain power. Consider that thousands of brilliant minds that are focused on trading problems and technology rather than on health research, building transportation infrastructure and pursuing other pressing social problems. While it is sensible for individual financial firms to seek the best talent, does Wall Street provide the most long-term benefit to society relative to alternative uses of this human capital? At a minimum, resources used in the human capital deployment decision do represent a significant opportunity cost to society.

On the other hand, it is easy to understate the benefits of trading because they are simply more difficult to readily isolate and evaluate. Trading in a free-market is essential to the production and flow of information, planning and managerial decision making, capital formation, risk transfer, etc. The strength of the economy depends on the resources devoted to the financial sector. We do know that capital allocation is a necessary component of a healthy economy; trading is an essential component of this process. Should not such essential activities command society's best minds and resources?

What is society's payback from this resource deployment? Turner (2010) argued that "There is no clear evidence that the growth in the scale and complexity of the financial system in the rich developed world over the last 20 to 30 years has driven increased growth or stability." If financialization did not significantly improve growth, stability, or firms' access to capital, how did it impact the economy? Turner suggests that we see little evidence, either at macro or individual investor levels that financial and economic risk has been substantially mitigated by increased market liquidity and trading. In fact, he suggests, speculation required to produce liquidity might have produced destabilizing and excessive momentum effects.

Furthermore, Godechot (2016), based on his study of 18 OECD countries, argues that the GDP share of the finance sector is a substantial driver of world wealth inequality, explaining between 20% and 40% of the increase in wealth inequality from 1980 to 2007. In particular, he argues that rise in financialization's contribution to inequality was driven largely by the increase in equities trading volume on national stock exchanges and by the large balances of shares held by banks.

Finance academics and professionals have traditionally accepted that trading, speculation and liquidity are beneficial to economies. This is surely true, at least to a point. Perhaps, increased levels of each are even more beneficial. Highly liquid financial markets reduce costs of capital, and, when functioning properly, they reduce uncertainty and improve information flows. But, do more liquidity and more activities that enhance liquidity always improve the economy? Clearly, there are benefits and costs to increased liquidity, but what are the appropriate trade-off levels between benefits and costs of additional financial activities? Has our economy reached the point of diminishing marginal returns to liquidity? Is there a point at which we should we consider transactions taxes discourage excessive devotion of resources to areas of economy that might need scaling back (see, for example, Summers & Summers, 1989)? Numerous countries all over the world, including the U.K., Australia, and China, have at some point at least temporarily imposed financial transactions taxes. The state of New York imposed a financial transactions tax in 1905 and retained it until 1981 after threats by the NYSE to move out of state.

Financialization is now an important research topic in economics. However, financialization is by no means a focus area of this book. Instead, we will focus on the theory and practice of financial trading from the perspective of the individual or institution. However, readings from the lists below should provide some insights on the important roles of financial trading in society.

Additional Reading

Harris (2003) is certainly one of the most useful introductory books on trading, especially with respect to microstructure. However, trading mechanisms are changing so rapidly that even this excellent book will be somewhat dated. Chapters 1 to 9 are much more comprehensive than the material presented here in Chapter 1, and very much worth reading. Kim (2007) provides excellent explanations of electronic and algorithmic trading technology in Chapters 1 to 3 and 6 to 7. Elton, Gruber, Brown, and Goetzmann (2010) provide a brief and readable introduction to securities markets in Chapters 2 and 3, and much of the remainder of the book will be useful to traders as well. Elton et al. is among the best MBA-level texts on portfolio analysis. The two articles by Klemperer (1999) and (2002) are excellent general introductory reviews of auction mechanisms. Stoll's 2003 review on market microstructure is an excellent overview of the topic, especially from an academic perspective. There are numerous businesses with websites providing information on brokerage services, data provision, trading platforms, and trading arcades. Zingales, in his 2015 address to the American Finance Association, raises the issue of financialization and the seeming divergence between academics' and the general public's perceptions of the net benefits to society of the practice of finance.

References

Arrighi, G. (1994). *The long twentieth century: Money, power, and the origins of our times*. London: Verso.

Battalio, R., Jennings, R., & McDonald, B. (2021). Deviations from time priority on the NYSE. *Journal of Financial Markets, 53*, 1–17.

Black, F. (1971). Towards a fully automated exchange, part I. *Financial Analysts Journal, 27*, 29–34.

Budish, E, Cramton, P, & Shim, J. (2015). The high frequency trading arms race: Frequent batch auctions as a market design response. *The Quarterly Journal of Economics*, *130*(4), 1547–1621.

Bureau of Economic Analysis, U.S. Department of Commerce. (2016). *Gross-domestic-product-(GDP)-by-industry data*. Available at <https://www.bea.gov/industry/gdpbyind_data.htm> Accessed 17.01.17.

Collin-Dufresne, P., & Daniel, K. (2014). *Liquidity and return reversals.* Working paper. New York: Columbia University.

Elton, E., Gruber, M., Brown, S., & Goetzmann, W. (2010). *Modern portfolio theory and investment analysis* (8th ed.). New York: Wiley.

Godechot, O. (2016). Financialization is marketization! A study of the respective impacts of various dimensions of financialization on the increase in global inequality. *Sociological Science*, *3*, 495–519.

Harris, L. (2003). *Trading and exchanges: Market microstructure for practitioners*. Oxford: Oxford University Press.

Kagel, J. H., Harstad, R. M., & Levin, D. (1987). Information impact and allocation rules in auctions with affiliated private values: A laboratory study. *Econometrica*, *55*, 1275–1304.

Kagel, J. H., & Levin, D. (1986). The winner's curse and public information in common value auctions. *American Economic Review.*, *56*(5), 894–920.

Kim, K. (2007). *Electronic and algorithmic trading technology: The complete guide*. New York: Academic Press.

Klemperer, P. (1999). Auction theory: A guide to the literature. *Journal of Economic Surveys*, *13*, 227–260.

Klemperer, P. (2002). What really matters in auction design. *Journal of Economics Perspectives*, *16*(1), 169–189.

Krippner, G. R. (2005). The financialization of the American economy. *Socio-Economic Review*, *3*, 173–208.

Kyle, A. S. (1985). Continuous auctions of insider trading. *Econometrica.*, *53*(6), 1315–1336.

Massa, A. (2019). Brokers profit from you even if they don't charge for trading. *Bloomberg Businessweek*, October 10. Available at <https://www.bloomberg.com/news/articles/2019-10-10/brokers-profit-from-you-even-if-they-don-t-charge-for-trading> Accessed 21.08.21.

O'Callahan, G. (1993). *The structure of the world gold market.* Occasional paper no. 105. Washington, DC: International Monetary Fund.

Philippon, T. (2013). Finance versus Wal-Mart: Why are financial services so expensive? In A. S. Blinder, A. W. Lo, & R. M. Solow (Eds.), *Rethinking the financial crisis* (pp. 235–246). New York: Russell Sage Foundation.

Philippon, T. (2015). Has the US finance industry become less efficient? On the theory and measurement of financial intermediation. *American Economic Review*, *105*(4), 1408–1438.

Silvennoinen, A., & Thorp, S. (2013). Financialization, crisis and commodity correlation dynamics. *Journal of International Financial Markets, Institutions and Money*, *24*, 42–65.

Stoll, H. R. (2003). *Market microstructure.* Working paper. Nashville: Vanderbilt University.

Summers, L. H., & Summers, V. P. (1989). When financial markets work too well: A cautious case for a securities transaction tax. *Journal of Financial Services Research*, *3*, 261–286.

Thompson, R. B., & Wright, A. L. (2005). *Equilibrium bidding strategies in common-value sealed-bid auctions.* Working paper. Tucson: Department of Mathematics, University of Arizona.

Turner, A. (2010). What do banks do? Why do credit booms and busts occur and what can public policy do about it? Chapter 1. In A. Turner, et al. (Eds.), *The future of finance: The LSE report* (pp. 5–86). London: London School of Economics and Political Science.

Vickrey, W. (1961). Counterspeculation, auctions, and sealed tenders. *Journal of Finance*, *16*, 8–37.

World Bank. (2021). *Stocks traded, turnover ratio of domestic shares (%)*. Available at <https://data.worldbank.org/indicator/CM.MKT.TRNR> Accessed 10.11.21.

Zingales, L. (2015). Presidential address: Does finance benefit society? *Journal of Finance*, *70*(4), 1327–1363.

1.9 EXERCISES

1. Is a futures contract a security?
2. How might long-term investors benefit from vigorous competition among short-term traders?

3. How does trading differ from speculating?
4. Suppose there was an open outcry English second-price auction. All bids are public. You are willing to bid a maximum of $100 for the auctioned item. You know that no one else will be willing to pay more than $200 for the object. The first bid by one of your competitors is for $10. What should you bid? Why?
5. An auction house has decided to sell a sculpture in an open outcry second-price auction. The auction house manager does not know the value any bidder will place on the work, but does know that every bidder will pay at least $500,000 (there is a gallery that has already announced its intent to pay at least this amount for the work, either from the gallery or from the successful bidder). Thus, the successful bidder can always sell the work for $500,000. In addition, the auction house knows that no one will be willing to pay more than $1,000,000 for the work because the only other identical sculpture (made from the same cast) is on the market for $1,000,000 and has not drawn any interest. Thus, the auction house believes that any particular bidder is willing to pay a random sum between $500,000 and $1,000,000 for the work (this random sum is the value of the work to the bidder), with an expected value of $750,000 and distributed uniformly (any value within this range is equally likely) between the minimum and maximum.
   a. Suppose that there are two bidders. What is the expected value of the second highest bid? Hint: If there are to be n independent draws from a uniform distribution on the range, $[x_1, x_2]$, $E[x_{\text{MAX}}] = x_1 + (x_2 - x_1)n/(n+1)$ and $E[x_{\text{MIN}}] = x_1 + (x_2 - x_1)/(n+1)$.
   b. Now, suppose instead that there are three bidders for the sculpture. What is the expected value of the second highest bid?
   c. Suppose that there are 999 bidders for the sculpture. What is the expected value of the second highest bid?
6. There are two ways to purchase T-Bills. The first is to enter a competitive bid at the auction where the bidding institution competes for a given dollar amount of the new issue based on how much it is willing to pay. Only the approximately 2000 securities brokers and dealers that are registered to operate in the government securities market are permitted to participate directly in the competitive bidding process. However, these registered firms may participate on behalf of their clients. Second, noncompetitive bids can be tendered by anyone where the prospective purchaser states how many bills he would like to purchase at the average price of accepted competitive bids. Competitive bidders for T-Bills generally enter their bids just before the deadline (1:00 PM Eastern Time) to participate in the auction. Noncompetitive bids are limited to $5 million and are normally due before 12:00 noon (Eastern Time) on the day of an auction. Noncompetitive bids are satisfied at the average price of successful competitive bids. The Treasury determines the dollar amount of competitive bids that it wishes to satisfy by subtracting the face values of the noncompetitive bids from the level of bills that the Treasury wishes to sell. Successful competitive bids are selected by ranking them, starting with the highest bid. Successful bidders obtain their bills at the prices that they bid; the lowest bid is referred to as the stop-out price. Consider the following example involving a Dutch auction of $20 billion in 91-day T-Bills,

where the bids (based on yields to maturity) by financial institution are given as follows:

Citigroup	$2.0 billion at 5.15%
Merrill Lynch	$4.5 billion at 5.20%
UBS	$5.5 billion at 5.25%
Deutsche Bank	$7.5 billion at 5.30%
JP Morgan/Chase	$5.5 billion at 5.35%
Bank of America	$6.5 billion at 5.40%

Further suppose that individual investors have placed noncompetitive bids totaling $2.5 billion. What is the bid-to-cover ratio in this auction? What is the stop-out price? Which bids will be satisfied?

7. **a.** In an all-pay auction, all bidders pay the amounts that they bid for the object to be auctioned, but only the winner, that is, whoever places the highest bid, takes possession of the object. This auction process has been likened to the system of political contributions, where all lobbyists make contributions to politicians, but only the highest-contributing lobbyists can exert influence. Similarly, the all pay auction might be similar to a ticket line at a limited-seating rock concert, where each bid is reflected by the amount of time that the bidder waits in line. Other all-pay auction scenarios might involve political campaign contributions, job promotions, and R&D races. Now, consider an all-pay auction for a $1 bill. Describe how such an auction structure might realize higher revenue for the seller of the $1 bill.

   b. Consider the entertainment shopping site, PennyCave.com (http://www.pennycave.com/; defunct as of June 2017). This site, as do other penny auction sites, auctioned electronic items such as music players and computer screens. Items were normally sold to the highest bidder at a price that is (or seems) substantially less than the value of the auctioned item. Bidders paid a fixed amount (e.g., $1) for each bid that they placed. The auction clock was set for a specific length of time (e.g., 24 hours), but a bid placed in the final seconds reset the clock for another 20 seconds. Its market structure follows (as of April 25, 2011):
   - The starting price on each auction is $0.00. Bid increments are $0.01.
   - Each bid costs the bidder one credit. Credits are purchased from PennyCave for costs ranging from $0.67 to $1, depending on the quantity purchased.
   - The price only goes up by $0.01 with each bid placed.
   - If a bid is placed in the final moments, the auction clock is reset for 20 seconds.
   - How does this auction structure compare to an all-pay auction structure? What is the advantage of having bid increments set at $0.01?

8. The Designated Market Maker (DMM) on the NYSE has received the following bids, each to purchase 100 shares of XYZ stock at the market open: 9.98, 9.99, 10.00, 10.01,

10.02, and 10.03. She has received the following offers, each to sell 100 shares of XYZ stock at the market open: 10.01, 10.02, 10.03, 10.04, 10.05, and 10.06.

a. Assume that this is the only market for the stock. What is the best bid prior to the market's open? What is the best offer?

b. Describe the demand curve or schedule for XYZ stock based on the data above. (Note: it is not continuous.)

c. Describe the supply curve or schedule for XYZ stock based on the data above. (Note: it is not continuous.)

d. What is the clearing price for this market? That is, at what price does supply equal demand for the stock?

e. Based on your solution to part d above, what bids execute at open? What offers execute at open?

f. Again, assume that this is the only market for the stock. What is the open NBBO after the initial clearing of orders?

9. a. How does a stop order differ from a limit order?

b. Suppose that a given market accepts only market and limit orders. Which of these two order types are more likely to be placed by buy-side market participants? Why?

c. Again, suppose that a given market accepts only market and limit orders. Which of these two order types are more likely to be placed by sell-side market participants? Why?

10. Exactly what does an investor expect from her broker when she places a stop limit order with a stop price to buy at 50 and a limit price of 50.10? Why might an investor place such an order?

11. High frequency trading is the practice of executing multiple transactions for securities followed by extremely short holding periods. Such market practices often require highly sophisticated computing and communications technologies. Many market observers have commented that high frequency trading (HFT) is unfair to small investors who lack the high-priced equipment and technology to compete with much larger traders. Many of these observers have proposed curbs on HFT activity, including having exchanges and other markets replace continuous trading with batch trading or call markets. Why would batch trading (call markets) reduce the level of HFT?

12. Suppose that orders arrive at the market in the following timed sequence, after the best-bid-offer (BBO) has been set at 49.98−50.02:

11:01.01 DMM bids 50.00 for 1,000 shares, setting the best bid
11:01.20 Floor broker bids 50.00 for 600 shares
11.01.40 Bid at 50.00 for 1,500 shares is placed in the Electronic Book
11:02.00 Bid at 50.00 for 1,000 shares is placed in the Electronic Book
11:02.10 Offer to sell 2,000 shares at market is placed in the Electronic Book

a. How would these orders be allocated for execution based on price/time priority?

b. How would these orders be allocated for execution on the NYSE?

13. Suppose that the last sale of Company X stock was at a price of $100.00. Further suppose that an investor wishes to place a market order to purchase 25,000 shares of Company X stock.

Market A		Market B	
# Shares	Offer ($)	# Shares	Offer ($)
30,000	100.00	10,000	100.00
40,000	100.02	10,000	100.01
10,000	100.05	10,000	100.03
20,000	100.06	20,000	100.04
30,000	100.07	40,000	100.05
10,000	100.09	40,000	100.05

Which of the two markets has greater depth?

14. Suppose that investing $10 million on eradicating mosquitoes carrying the Zika or other infectious virus will have a clear and significant impact on the spread of this disease, saving over 500 lives and preventing disabling conditions in another 500 persons. Suppose that an equal investment into a new trading platform will produce $300 million in additional revenues to a brokerage firm. Consider the following rhetorical questions:

a. Which investment is likely to produce more social benefits? Which investment is likely to produce more private benefits to the decision-maker?

b. Would an institution that makes the investment that produces more social benefit be more highly regarded by society than the institution making the investment with higher private benefits?

c. Should governments play a significant role in making such decisions?

CHAPTER

2

Securities Markets

2.1 EXCHANGES AND FLOOR MARKETS

The Securities and Exchange Act of 1934, in a manner somewhat different from our definitions in the previous chapter of this book, defined an *exchange* to be:

> any organization, association, or group of persons, whether incorporated or unincorporated, which constitutes, maintains, or provides a market place or facilities for bringing together purchasers and sellers of securities or for otherwise performing with respect to securities the functions commonly performed by a stock exchange as that term is generally understood, and includes the market place and the market facilities maintained by such exchange.

An exchange is a physical or virtual meeting place drawing together brokers, dealers, and traders to facilitate the buying and selling of securities. Thus, exchanges include the floor-based markets where brokers and traders physically congregate as well as many virtual meeting sites and screen-based systems provided by *electronic communications networks* (*ECNs*). In the United States and many other countries, exchange transactions are executed through some type of auction process. Exchanges in the United States are intended to provide for orderly, liquid, and *continuous markets* for the securities they trade. A *continuous market* provides for transactions that can be executed at any time for a price that might be expected to differ little from the prior transaction price for the same security, whereas a *call market* allows for trading of securities only at specific points or intervals during the trading day. In addition, exchanges traditionally served as *self-regulatory organizations* (SROs) for their members, regulating and policing their behavior with respect to a variety of rules and requirements. However, in the United States, the *Financial Industry Regulatory Authority* (*FINRA*) has provided for self-regulation of the two major U.S. stock markets since the 2007 consolidation of the *National Association of Security Dealers* (NASD) and the Member Regulation, Enforcement and Arbitration operations of the New York Stock Exchange.

DOI: https://doi.org/10.1016/B978-0-323-90955-6.00002-1

© 2023 Elsevier Inc. All rights reserved.

2.2 THE WAY IT WAS

National and Regional Exchanges

Understanding the history and evolution of securities markets is important for understanding their current structures and practices. U.S. stock markets have been the world's most significant and most followed, and merit particular attention. Traditionally, the New York Stock Exchange (NYSE), now a unit of the Intercontinental Exchange Group, and the American Exchange (ASE or AMEX) were regarded as the two national exchanges in the U.S., but the NYSE acquired the ASE in 2008. The ASE's successor's market is now known as NYSE MKT, which trades small cap equities, with physical floor facilities alongside the NYSE in New York. Until the 1990s, numerous "regional exchanges" continued to operate. Regional exchanges, a term whose meaning is now somewhat in flux, traditionally referred to all but the two national exchanges, including those in Chicago (formerly the Midwest Exchange), Boston, and Philadelphia, the descendant of the Philadelphia-Baltimore-Washington Exchange. The Pacific Exchange, which originated as two separate regional exchanges in San Francisco and Los Angeles, maintained separate trading floors in San Francisco and Los Angeles, and was eventually acquired by the Archipelago Group. Archipelago has since merged (in 2006) with the New York Stock Exchange to form the New York Stock Exchange Group which, as part of the NYSE Euronext Group in 2013, was acquired by the Intercontinental Exchange (ICE), which then spun off Euronext and remains the owner of the NYSE.

Regional exchanges traditionally listed stocks that would not qualify for national exchange listing as well as provided local or competing markets for a number of national exchange listed securities. However, many regional exchanges no longer fit this mold, and most do not consider themselves to be regional. In some instances, either lower transactions costs or better prices can be obtained for clients for dual-listed securities on the regional or smaller exchanges; to some extent, competition provided by these regional exchanges may result in lower transactions costs for investors on national exchanges. In addition, different market structures and rules might provide for unique execution opportunities for clients and investors. There are also a number of exchanges for other securities such as the Chicago Board of Trade (CBT—for futures contracts), The Chicago Board Options Exchange (CBOE), the Kansas City Board of Trade, and the Commodities Exchange (COMEX). Finally, there are several exchanges without actual trading floors, including the National Exchange (formerly the Cincinnati Exchange, currently located in Chicago) and the International Securities Exchange (ISE). These exchanges operations are based entirely on electronic platforms. Table 2.1 provides a more comprehensive listing of exchanges as recognized by the SEC as of April 2021 along with several of the more significant U.S. futures exchanges (Designated Contract Markets) registered with the CFTC.[1]

Early Ancestors to Modern Exchanges

Until recently, U.S. stock markets and many others had evolved rather slowly over the centuries, hanging on to their roots, traditions, rules, and regulations. Much of their

[1]These markets have registered with the SEC as exchanges, pursuant to Article 6 of the Securities Exchange Act of 1934 according to the Securities and Exchange Commission (2021b).

TABLE 2.1 U.S. Stock and Futures Exchanges as of April 2021

U.S. Stock and Options Exchanges
NYSE American LLC (formerly NYSE Amex LLC and the American Stock Exchange)
Cboe BZX Exchange, Inc. (formerly BATS Exchange, Inc.)
Cboe BYX Exchange, Inc. (formerly BATS Y-Exchange, Inc.)
BOX Exchange LLC
Nasdaq BX, Inc. (formerly the Boston Stock Exchange)
Cboe C2 Exchange, Incorporated
Cboe Exchange, Inc. (formerly Chicago Board Options Exchange, Incorporated)
NYSE Chicago, Inc. (formerly Chicago Stock Exchange, Inc.)
Cboe EDGA Exchange, Inc. (formerly Bats EDGA Exchange, Inc.)
Cboe EDGX Exchange, Inc. (formerly Bats EDGX Exchange, Inc.)
Nasdaq ISE, LLC (formerly International Securities Exchange, LLC)
The Investors Exchange, LLC
Long-Term Stock Exchange, Inc.
MEMX LLC
Nasdaq GEMX, LLC (formerly ISE Gemini)
Nasdaq MRX, LLC (formerly ISE Mercury)
Miami International Securities Exchange
MIAX Emerald, LLC
MIAX Pearl, LLC
The Nasdaq Stock Market LLC
NYSE National (formerly National Stock Exchange, Inc.)
New York Stock Exchange LLC
NYSE Arca, Inc.
Nasdaq PHLX, LLC (formerly Philadelphia Stock Exchange)
U.S. Futures Exchanges (Partial Listing)
Chicago Board of Trade
CBOE Futures Exchange, LLC
Chicago Mercantile Exchange
OneChicago, LLC
Intercontinental Exchange
NQLX LLC

current structures and idiosyncrasies still draw from this evolution. Understanding this evolution is helpful to understanding their current structures, policies, and procedures. Precursors to modern stock exchanges might have existed in Egypt as early as the 11th century, where it is believed that Jewish and Islamic brokers traded a variety of credit-related instruments. Thirteenth-century Bruges (Belgium) commodity traders assembled in the van der Beurse family home (and inn), ultimately becoming the "Bruges Beurse." Additional bourses opened elsewhere in Flanders and Amsterdam, generally trading alongside merchants of commodities. Codifying of and enforcing trading rules and regulations by the exchange in Antwerp and later exchanges substantially facilitated transactions and reduced transactions costs among the traders from many countries, perhaps even contributing to the slow decline of the important merchant guilds over subsequent centuries (see Raj, 2017). The Amsterdam Stock Exchange opened in the early 17th century, trading shares of the Dutch East India Company. The exchange continues to operate as a unit of Euronext, a family of European exchanges, and is the world's longest continuously operated exchange. Several early exchanges began in coffee houses and taverns, where brokers and dealers would meet to trade securities. For example, the London Stock Exchange started operations in about 1698 in Jonathan's Coffee House on Change Alley.

The first securities exchange to operate in the United States was in Philadelphia, the second, in New York, which opened operations in 1791. The New York Stock Exchange began operations outdoors after the May 17, 1792 signing of the "Buttonwood Agreement." This contract, signed by 24 brokers under a buttonwood tree at what is now 68 Wall Street, fixed brokerage commissions among its signatories and prohibited off-board trading (i.e., traders and brokers were to trade only with each other). In earlier days, exchanges often operated outdoors so that brokers could call out their orders from their office windows to the street where transactions actually took place. In late 1793, this group of brokers obtained the Tontine Coffee House, which was used for cold weather trading until the 1880s. In 1865, the New York Stock Exchange adopted its current name. The American Stock Exchange, known as the New York Curb Exchange until 1953, did not move indoors until June 27, 1921. For much of its history, NASDAQ was not considered to be an exchange because it did not have a physical trading floor. Founded in 1971 by the National Association of Securities Dealers (NASD), NASDAQ—formerly known as the National Association of Security Dealers Automated Quotation System—has served as an electronic stock market allowing brokers and dealers to efficiently transmit bid, offer, and close prices for securities.

Traditional New York Stock Exchange Structure

Since the New York Stock Exchange is the largest and arguably the most important exchange in the world, we will first emphasize its structure and organization, first past and then present. Until 2006, the NYSE was a sort of hybrid corporation/partnership whose members faced unlimited liability.[2] At the same time, the NYSE structure was unaffected by the death or departure of any one of its 1366 members. Only members who owned or leased seats had trading privileges and there were four types of members,

[2]This type of organizational structure for exchanges, quite common until the 1990s, is becoming increasingly rare.

discussed below. Each member was an individual and most were sponsored by or associated with a brokerage firm or other organization (typically called a member firm). A membership was referred to as a seat, which was purchased (for as much as $4 million in 2005, and last sold on December 29, 2005 for $3,505,000) through a competitive market process. Full membership was limited to 1366 seats, although there had been special limited trading permits issued as well. Now, under the new structure, one can join the 1366 members by purchasing a full trading license that permits the member to trade for one year. As of January 2017, the annual fee for a full trading license was $50,000, billed monthly. The four types of traditional memberships were as follows:

House Broker: Also known as a commission broker, executed orders on behalf of clients submitting orders through brokerage firms. Until 2006, approximately half of the members on the floor were commission brokers. This and other broker roles have been taken over by "trading floor brokers."

Independent Broker: Also called a two-dollar broker, executed orders on behalf of commission brokers when activity was high. Alternatively, sometimes commission brokers executing orders on behalf of off-floor clients were referred to as floor brokers. This type of distinct membership no longer exists.

Floor Trader: Executes orders on her own trading accounts. As of 1992, there were approximately 40 floor traders on the NYSE and this number continued to decline through 2006. However, the NYSE has since created the "supplemental liquidity provider" role, which is intended to enhance market liquidity by allowing for proprietary trading. We will discuss this new role shortly.

Specialist: Was charged with the responsibility for maintaining a continuous, liquid, orderly market for the securities in which he specialized. He also kept the book (records) for unexecuted transactions and bought and sold on his own account. There were approximately 400 specialists on the NYSE in 1992. Specialists were typically selected on the basis of their performance as specialists for other securities and other qualifications. Specialists were involved in approximately 20% of NYSE transactions as a principal. Because of their dual roles in running the market for securities on behalf of the exchange and as traders buying and selling those securities at a profit, the specialist's role was particularly interesting and somewhat controversial. The specialist was expected to fulfill his obligation to maintain a continuous, orderly, and liquid market; this role may have conflicted with the desire to earn trading profits. The specialist has been replaced by the "designated market maker" (DMM), to be discussed shortly.

Traditionally, trading activity for the shares of approximately 2779 listed firms (prior to 2006) centered around trading posts located around the floor throughout various rooms. Brokers and traders would congregate about these posts and specialists who oversee trading activity. Clerks employed by the exchanges recorded trading activities and relayed records to systems that disseminated data to the investing public. However, this system has mostly given way to automated screen-based trading, DMMs who are expected to maintain liquid and orderly markets for specific stocks, and off-floor supplemental liquidity providers who are paid fees by the exchange to execute transactions.

Transforming the NYSE

On April 20, 2005, the New York Stock Exchange dropped a bombshell on the investing public, announcing its intent to be taken over by Archipelago, the successor to the Pacific Stock Exchange. In a single sweep, the NYSE would be merged, privatized, expanded into derivatives markets, and have a springboard for its transformation to an electronic trading market. Stock markets have traditionally emphasized roles of the specialists (replaced by DMMs) who conduct auction markets and maintain liquidity for off-floor brokers and traders. Since the founding of the NYSE in the late 1700s, exchange markets have brought together widely dispersed brokers, dealers, and individual investors, creating highly organized, visible, and regulated environments intended to facilitate trading and price discovery. The traditional floor-based stock exchange in the United States centered around a specialist with a regular "trading crowd" that focused its energies on a small number of stocks. The specialist conducted an auction market for the trading crowd that gathered about. Floor brokers entered the trading crowd with orders initiated by larger off-floor traders, while smaller and routine orders were more likely to be routed to the specialist through an electronic routing system such as the now-obsolete *SuperDOT*, the "super designated order turnaround system."

The merger with Archipelago and the embrace of modern technology have completely changed the way that the NYSE conducts its markets. Now, the roles of the specialist have been passed on to *designated market makers* (DMMs) whose responsibilities and rights differ somewhat from their specialist predecessors. DMMs have no trading advantages relative to other market participants, have no special access to order books, and are compensated by listing firms to provide liquidity. The DMM manually manages the market open and close for her securities, overseeing the process of determining the market clearing prices. The exchange's matching engine is a software algorithm that continuously matches buy and sell orders for execution through the trading day, though the process is overseen by the DMM, who commits capital to maintain liquidity in her market. In addition, the DMM steps in with her own capital during period of market turbulence, seeking to maintain a balance between buy and sell orders, maintain price continuity and to reduce market volatility. We will further discuss the DMM roles shortly.

The NYSE has overhauled other technologies as well. For example, in July 2009, the SuperDOT system was replaced by the NYSE *super display book system* (SDBK) for routing, execution and reporting orders. This new system enables customers to execute orders in as quickly as five milliseconds without broker intervention.

Sophisticated traders are very concerned with the development of fast and efficient buy-and-sell order processing systems, order executing (clearing and settling) systems, and systems for the analysis of investments. Technology affects order processing, routing, and matching of trades, as well as valuation, hedging, and portfolio risk management techniques. The NYSE has responded with numerous enhancements to facilitate trading activity. For example, it has dramatically increased its ability to handle increasing transaction volume (typically averaging between one and two billion equity shares between mid-2016 and mid-2017) due to improved technology. For example, the NYSE's ECN, the SDBK permits orders to be routed electronically by brokerage firms for execution. Most program trades are also submitted through the SDBK system, accounting for over 10% of all orders. Overall, the SDBK system

executes approximately 99% of all NYSE orders. For these SDBK orders, the brokerage firm need not have the order presented at the trading post by a commission broker. This means that the order can be executed faster, perhaps within 5 milliseconds. Equities orders that are executed by a broker on the floor itself are much larger, averaging approximately 29,000 shares. When the NYSE opens at 9:30 EST, orders that have accumulated matched and the DMM attempts to establish a market clearing price. The DMM will attempt to open the security at a price that deviates as little as possible from the prior day's close at 4:00 PM.

Demutualization and Governance Changes

A mutual organization is owned by members, typically employees, clients, or suppliers. Mutual stock exchanges are owned by members who have trading rights in the exchange's market. Accompanying the improvement of exchange technology in trading has been a trend to demutualize exchanges. This demutualization trend has meant that many exchanges have restructured to become stockholder-owned corporations. Euronext and the Australian, Pacific, and London Exchanges were among the first to demutualize in 1997 through 1999, following the Stockholm Exchange in 1993. NASDAQ demutualized in 2000 and the NYSE in 2006. Most exchanges justified restructuring to for-profit institutions based on their needs to raise capital to finance their technology and infrastructure expenses. Exchange demutualization has also provided for more intense competition among markets. Exchange self-regulation and other regulatory issues remain important concerns for these restructured institutions. To avoid conflicts of interest, the NYSE Group created a not-for-profit corporation, NYSE Regulation, Inc., with the New York Stock Exchange LLC as its sole equity member. This new corporation performs all the SRO (self-regulatory organization, to be discussed later) regulatory responsibilities that had been conducted by the NYSE. NYSE Regulation seeks to ensure that listed companies comply with NYSE listing regulations, actively engages in the exchange rule-making process, and monitors exchange trading activity for suspicious behavior.

The NYSE merger followed on the heels of the Richard Grasso compensation quagmire, where former CEO Grasso and the NYSE were heavily criticized for his $180-million pay package. Grasso resigned in 2003 and was followed in several months by John Thain who approached Archipelago about merging. The $6 billion merger deal granted the 1366 NYSE members about 64% of the combined firm's shares and the remainder to Archipelago shareholders (30%), NYSE top management (5%), and lower-level managers and employees (1%). Interestingly, the 5% granted to NYSE CEO John Thain and other top managers was valued at approximately $300 million, more than the disputed sums paid to former CEO Richard Grasso. A number of NYSE member-seat holders, notably Kenneth Langone, former chairman of the NYSE Board of Directors, complained about the price accepted by NYSE management. However, a seat on the NYSE sold for $960,000 in January 2005, and $1.6 million on April 20, 2005, shortly before the takeover announcement, and for $3 million in July 2005, shortly after the takeover announcement. By year-end, the price of a seat had risen to $4 million before dropping to $3.505 million at year-end. In addition, some NYSE members complained about Thain's former employer, Goldman Sachs, representing both sides in the deal. Furthermore, Goldman Sachs owned several seats on the NYSE and maintained

substantial holdings of Archipelago shares. However, two other investment banks, Greenhill and Lazard, did required due diligence for the deal. The deal was valued by these institutions at between $3 billion and $5 billion, a substantial discount below the value implied by the July 2005 market price of Archipelago shares.

Diminished Floor Trading Activity

As the NYSE developed, acquired, and improved its ability to handle electronic transactions, actual in-person floor trading activity diminished. As of the end of 2006, the NYSE trading floor exceeded 43,000 square feet. During 2007, the NYSE closed trading rooms; by November 2007, slightly more than half of this trading space was available for in-person transactions. In fact, as early as 2007, trading activity on the NYSE trading floor produced less than 10% of NYSE revenue. More important sources of revenue to the NYSE now include listing fees charged to firms that want their shares traded on the NYSE and from selling price, quotations, and other transactions data to the public. The trading floor crowd (including members, clerks, etc.) has diminished in numbers from over 3000 to less than 1700. Less than half of NYSE volume is executed on the floor, and DMMs (specialists) have been involved in less than 1 in 30 transactions as opposed to 1 in 7 as recently as 2002.

The NYSE functioned without any floor trading when the COVID-19 pandemic forced the exchange to close its physical trading floor from March 23, 2020, until it gradually reopened beginning May 26, 2020. The suspension of floor trading meant that *D-orders* (discretionary orders), which can be executed only by floor brokers, were also suspended, having a significant impact on daily closing auctions, the period most impacted by direct human trading. Electronic trading continued throughout this period while floor employees worked remotely. Closing the floor did reduce both liquidity and pricing efficiency, even relative to Nasdaq, which was restored by reopening the floor (Chung & Chuwonganant, 2021). Through the broader period extending from February to August 2020, NYSE stock markets experienced unprecedented volatility largely due to the impact of the pandemic on the economy.

Current New York Stock Exchange Equities Membership Types

As discussed above, the specialist role yielded to that of the DMM, who commits capital to support liquidity and maintains critical management roles in auctions for securities. The specialist's liquidity providing function was taken over by both the DMM and *supplemental liquidity providers* (SLPs), outside firms that commit to maintain liquidity in specific stocks. The DMM has the obligation to maintain fair and orderly markets and facilitate price discovery in his assigned securities. The DMM opens and closes the market for its securities each day, using his own capital (or that of his employer) to resolve order imbalances and to seek market clearing prices (prices at which demand equals supply) for the securities for which he is responsible. The DMM is required to quote at the national best bid or offer (NBBO) a specified percentage of the time (typically 40%), provide price improvement (that is, facilitate higher bids to accommodate prospective sellers and lower offers to accommodate prospective buyers), and match incoming orders based on a preprogrammed capital commitment schedule. DMMs for a listed security can be selected by the issuer of the listed

security or the issuer can delegate authority to the NYSE to select an eligible DMM. NYSE DMM firms include or have included BofA Merrill, Barclays Capital, GTS Securities, GETCO Securities LLC, Citadel Securities, and Spear, Leeds & Kellogg Specialists LLC.

A second type of NYSE equities membership is the *trading floor broker* who works orders on behalf of clients on an agency basis. Trading floor brokers are positioned at the point of sale on the exchange floor during openings, closings, and unique intraday occurrences to execute trades.

The third type of NYSE equities membership is the SLP, a high-volume member trading electronically off of her own proprietary account adding liquidity to the NYSE. SLP members are rewarded, often with rebate fees paid by the NYSE (averaging $.0019 in July 2016), for aggressively providing liquidity and improving markets. They are expected to maintain a bid or offer at the NBBO in each assigned security during at least 10% of the trading day. Major NYSE supplemental liquidity providing firms have included Barclays Capital, Citadel Securities LLC, Goldman, Sachs & Company, HRT Financial LLC, and KCG Americas LLC.

Option Exchanges

Most stock options in the United States are traded on one of the options exchanges. The oldest stock options exchange is the Chicago Board Options Exchange, which also owns the C2 Options Exchange. The NYSE has grown in stature in options markets with the NYSE Amex and NYSE Arca, acquired through takeovers. NASDAQ OMX operates the NSDQ and PHLX (formerly the Philadelphia Exchange). In addition, the Boston Options Exchange (BOX, which is jointly owned by the TMX Group and a consortium of broker dealers) and ISE (owned by Nasdaq) both maintain stock option markets. Table 2.2 provides volume data for all of these and other equity options exchanges. The Options Clearing Corporation (OCC), jointly owned by the U.S. options exchanges, is technically the counterparty for all listed options transactions in the United States. This means that all option buyers own options that were written by the OCC and all option writers are obligated to the OCC. This arrangement, given the financial stability of the OCC and each of the exchanges that own and back its obligations and all of the brokerage firms that back the obligations of the exchanges have effectively eliminated default risk in listed options trading.

Options trading took place entirely over the counter until the 1973 opening of the Chicago Board Options Exchange (CBOE). The nature of options trading lends itself to technological development at every stage of the trading process. First, because of option-trading strategies' emphases on relative valuation and stochastic processes, the valuation and portfolio analyses require mathematical analyses that are computer-based. Maintenance of these strategies requires abilities to rapidly trade in and out of multiple positions simultaneously, relying on electronic communication technologies capable of handling large amounts of data instantaneously. Technologies offered through the Internet offer these capabilities, forcing exchanges to anticipate and quickly improve on developments in order to compete.

Competition among options exchanges continues to intensify, fostered by a myriad of new technologies and the introduction of equity option multiple listings beginning in the fall of 1999. The International Securities Exchange was launched in May 2000 as the first fully electronic exchange in the United States, adding a significant dimension of competition to the then largely floor-based markets offered by the CBOE, AMEX, and PHLX. In addition,

TABLE 2.2 U.S. Equity Options Exchanges (Data for Calendar Year 2020)

Exchange	Volume	Daily Volume	% of Market Share
AMEX	578,931,703	1,590,472	8.27
ARCA	711,687,063	1,955,184	10.16
BATS	686,855,135	1,886,965	9.81
BOX	202,168,886	555,409	2.89
C2	228,581,695	627,972	3.26
CBOE	893,034,904	2,453,393	12.75
EDGX	295,857,198	812,795	4.22
EMLD	203,740,528	559,727	2.91
GEM	392,570,795	1,078,491	5.60
ISE	543,855,569	1,494,109	7.76
MCRY	52,111,029	143,162	0.74
MIAX	330,036,680	906,694	4.71
MPRL	293,410,388	806,072	4.19
NOBO	14,846,436	40,787	0.21
NSDQ	688,942,003	1,892,698	9.84
PHLX	887,674,136	2,438,665	12.67
OCC Totals	7,004,304,148	19,242,594	100

Source: The Options Clearing Corporation (2021).

the ISE traded options with well-established markets on the other exchanges. In addition to the ISE, well-established foreign exchanges such as Eurex electronically trade contracts on U.S. equity instruments, further intensifying competitive pressures on major U.S. options exchanges. In fact, practically all European and Asian securities and derivatives exchanges have become entirely or almost entirely electronic. This increased competition has led to a number of important results, including narrowing spreads (8% between 1999 and 2000 alone according to a Securities and Exchange Commission (SEC) study; also see Battalio, Hatch, & Jennings, 2004), highly uncertain futures for competitors, and the development and implementation of innovative quotations and trading systems.

Options exchanges have innovated substantial technological advances to maintain and even anticipate developments in other markets. For example, exchange-initiated technological advances include NASDAQ's automated exchange, *SuperMontage*. Options exchanges have experienced launching of the Chicago Board Options Exchange CBOE Command, launched in 2012 as CBOE's trade platform and engine located in Secaucus New Jersey. Such trading systems provide for instantaneous filling and confirmation of orders at the best prices along with transparency of trades. Asset managers rely on such systems to obtain speedy access to information, to route and execute orders, and to obtain trade

confirmations. Several exchanges are providing for off-site market makers such as the CBOE *Remote Market Makers* (RMMs). These systems facilitate trading by off-floor investors and enhance liquidity by expanding the membership of market makers. They offer market makers direct access to the trading floor, enabling them to participate in the provision of market liquidity and subjecting them to the same market-enhancing responsibilities without requiring their physical presence.

2.3 OVER-THE-COUNTER MARKETS AND ALTERNATIVE TRADING SYSTEMS

During most of the 20th century, the two national exchanges (NYSE and AMEX) along with a small number of regional exchanges hosted all U.S. exchange-based trading. All other trading was conducted either directly between counterparties or intermediated by broker dealers in what was called the *over-the-counter (OTC)* markets. The creation of the electronic bulletin board NASDAQ by NASD, followed by its evolution into a full-fledged ECN, and then its recognition as an exchange, along with the introduction of numerous alternative trading systems such as BRUT and BATS, blurred the distinction between exchange and OTC markets.

Over-the-Counter Markets

The *OTC markets* have traditionally been defined as nonexchange markets. Where the expression applies, OTC markets are made up of broker-dealer houses that execute transactions on behalf of client accounts as well as their own. In addition, included are transactions executed directly by trading counterparties. Almost all federal, municipal, and corporate bonds are traded in the OTC markets as well as many derivative products, structured products, and shares of smaller corporations.

Securities brokers and dealers in the OTC markets are members of FINRA, the Financial Industry Regulatory Authority, which absorbed the National Association of Securities Dealers (NASD). FINRA is an independent, self-regulated agency cooperating with the SEC. As of March 2017, 3819 broker/dealer firms and 634,874 registered representatives were regulated by FINRA. FINRA also operates the largest securities dispute resolution forum in the world, closing 3635 arbitrations and 172 mediations in 2016.

OTC markets for equities are comprised of a number of venues. For example, the Over-the-Counter Bulletin Board (OTCBB) is an interdealer electronic quotation system that displays real-time quotes, last-sale prices, and volume information for many OTC equity securities. OTCBB is not an exchange; securities are not actually traded through the system. OTCBB securities include national, regional, and foreign equity issues, warrants, unit investment trust issues (UITs), American depositary receipts (ADRs), and direct participation programs issues (DPPs). Investors do not have direct access to OTCBB.

Another important OTC equity venue is the OTC Markets Group, Inc., informally known as “Pink Sheets.” This group also facilitates trading of shares of nonlisted companies, providing quotation, messaging, and information services to broker-dealers. Pink Sheets maintains categories based on the level of financial and corporate disclosure

provided by companies quoted through their systems. For example, shares quoted in OTCQX, the top tier of the OTC market, meet certain financial standards, and are subject to qualitative reviews. OTCQB tier companies (mid-tier) also provide significant information disclosure to the marketplace. Pink is the third tier of the OTC market and is further categorized into three tiers characterized by their information disclosure requirements, which might be virtually nonexistent in the lowest tier.

The OTC markets also include the so-called third and fourth markets. Third markets exist for listed securities traded between brokers off any exchange. Fourth markets are those where institutions trade securities directly among themselves in the so-called *upstairs markets*. These markets are particularly important for bank-to-bank derivatives and structured products transactions. Institutions have used these markets for block trades (transactions involving more than 10,000 shares or $200,000 of a security), particularly when such a transaction might pose liquidity problems on an exchange. The upstairs markets might account for as much as half of all stock transactions volume, although statistics are difficult to estimate and confirm.

Stock exchanges and other markets vary with respect to how orders are intermediated. For example, the NYSE permits many of its members to act as both agents (brokers) and dealers. This means that an order counterparty might be either a dealer trading on her own account or a broker representing a client. On the other hand, the OTC markets are usually dealer markets, where investors transact only with or through dealers.

Alternative Trading Systems

An ATS might be defined as a securities trading venue that is not registered with the SEC as an exchange. Alternative trading systems provide electronic forums for linking dealers with each other and with institutional investors. They also match and cross trades in the upstairs market (fourth market). We will discuss the more precise SEC definition shortly. The four primary types of ATSs are:

1. *Electronic Communication Networks (ECNs)*, which are virtual meeting places and screen-based systems for trading securities. These ECNs include or have included Bloomberg Tradebook (which bills itself as a more comprehensive full-service agency broker), CBSX, and Direct Edge. Some ECNs have been acquired by or are now exchanges.
2. *Dark Pools and "Crossing Networks,"* where quotations for share blocks are matched anonymously. These ATSs include or have included Credit Suisse Crossfinder, Goldman Sachs Sigma X2, Liquidnet, Posit (ITGP), MS Pool (Morgan Stanley), and CITIBLOC (Citigroup). Participants in crossing markets often enjoy reduced transactions costs and anonymity but often must wait for counterparty orders to accumulate before transactions can be executed.
3. *Internalization Crossing Networks* such as UBS PIN that allows brokers or firms to fill orders from that firm's own internal supply of stock.
4. *Voice-Brokered Third-Party Matching*.

A small listing of alternative trading systems is provided in Table 2.3. As of October 2021, the SEC listed a total of 85 Alternative Trading Systems with Form ATS on file (Securities and Exchange Commission, 2021a).

TABLE 2.3 Sampling of Alternative Trading Systems

Name	Instruments	Features
AQUA	Equities	Block trading
Global OTC	Equities	Affiliated with Archipelago
Bloomberg Tradebook	Equities, FX, and derivatives	Bills itself as an integrated full-service agency broker
CITIBLOC (Citigroup)	Equities	Internal Crossing Network
Crossfinder (Credit Suisse)	Equities	Bills itself as the world's largest dark pool; internal crossing network
Level ATS	Equities	Dark pool; crossing network
Liquidnet	Equities	Block negotiation (crossing)
Sigma X (Goldman Sachs)	Equities	Crossing network owned by Goldman Sachs
UBS ATS	Equities	Crossing network

The most important of the ECNs dealing with equities have included *BRUT* (purchased by NASDAQ) *Instinet*, *REDIBook*, and *Archipelago*, with the latter two having merged in 2002. Instinet, the first of the ECNs, was founded in 1967 as Institutional Networks, became a wholly owned subsidiary of Reuters in 1987, took its shares through an IPO in 2001 and merged with Island shortly afterwards before going private again in 2005. The focus of Instinet's activities was block trading (10,000 or more shares), seeking to provide its clients with a high level of confidentiality. Major ECNs dealing with foreign exchange include Hotspot and Currenex.

The SEC defines an ATS to be an automated system that centralizes, displays, crosses, matches, or otherwise executes trading interest, but is not currently registered with the Commission as a national security exchange or operated by a registered securities association. However, the SEC had defined the term "exchange" to include "any organization, association, or group of persons that: (1) Brings together the orders of multiple buyers and sellers; and (2) uses established nondiscretionary methods (whether by providing a trading facility or by setting rules) under which such orders interact with each other, and the buyers and sellers entering such orders agree to the terms of a trade." Under this new definition, at least several ATSs might be considered exchanges, even if they are not the traditional "brick-and-mortar" type. In the past few years, several large ATSs, including BATS and DirectEdge, have registered with the SEC as exchanges. Thus, consider the following definition provided by Rule 300(a) of SEC Regulation ATS:

> Alternative trading system means any organization, association, person, group of persons, or system:
>
> 1. That constitutes, maintains, or provides a market place or facilities for bringing together purchasers and sellers of securities or for otherwise performing with respect to securities the functions commonly performed by a stock exchange within the meaning of Rule 3b-16 under the Securities Exchange Act of 1934; and

2. That does not:
   i. Set rules governing the conduct of subscribers other than the conduct of such subscribers' trading on such organization, association, person, group of persons, or system; or
   ii. Discipline subscribers other than by exclusion from trading.

What the ATS does *not* do is what distinguishes it from an exchange. However, these sorts of distinctions might remain in flux elsewhere.

The NYSE share of NYSE-listed trade executions have been steadily dropping as electronic exchanges and ATSs have grown in importance. For example, NYSE's share of trading in its own listed stocks declined to 67.3% by December 2006, then down further to 22.5% by December 2013 (Baldauf & Mollner, 2020; NYSE Euronext, 2014; Weber, 2007). Table 2.4 provides a listing of important U.S. equity markets along with their shares of daily volume as of January 2021. Notice towards the bottom of Table 2.4 that FINRA and

TABLE 2.4 Equity Trading Center Volume, January 2021

Market Participant	Total Shares	Dollar Value Traded	% of Market
BYX Equities (Y)	4,137,801,358	$164,362,425,858	1.39
BZX Equities (Z)	12,769,337,168	$645,067,806,443	5.47
EDGA Equities (J)	3,480,732,058	$174,945,683,289	1.48
EDGX Equities (K)	22,962,628,303	$760,490,312,283	6.45
Investors Exchange (V)	5,625,233,619	$376,293,561,467	3.19
LTSE (L)	69,475	$10,676,834	0.00
MIAX Pearl (H)	506,156,135	$26,956,007,448	0.23
Members Exchange (U)	1,895,723,491	$39,794,283,895	0.34
NASDAQ (Q)	44,340,422,037	$2,203,888,914,146	18.68
NASDAQ BX (B)	1,971,827,519	$99,352,734,306	0.84
NASDAQ PSX (X)	1,770,154,381	$84,251,093,091	0.71
NYSE (N)	24,875,708,751	$1,179,771,087,659	10.00
NYSE American (A)	1,580,641,180	$32,221,094,288	0.27
NYSE Arca (P)	27,441,280,256	$1,117,494,076,355	9.47
NYSE Chicago (M)	648,487,597	$73,538,097,858	0.62
NYSE National (C)	2,322,010,586	$69,706,646,111	0.59
FINRA / NYSE TRF (DN)	33,365,957,723	$775,963,781,304	6.58
FINRA / Nasdaq TRF Carteret	105,497,703,899	$3,941,105,469,060	33.41
FINRA / Nasdaq TRF Chicago	812,442,085	$30,083,918,819	0.26
Totals	296,004,317,621	$11,795,297,670,514	100

Data reflects total daily volume in exchange-listed securities for the month of January 2021.
Source: Cboe, U.S. Equities Market Volume Summary (2021): https://www.cboe.com/us/equities/market_share/.

TRF (Trade Reporting Facility) volume for nonexchange trading in exchange listed shares totals represent over 40% of total exchange-listed volume.

Since the 1990s, equity markets have become less concentrated as the major exchanges have lost their monopolies. The newer, more *fragmented* markets enable investors to route their orders to a number of different centers that compete for order flow. Whereas concentrated markets tend to provide for enhanced price transparency, reduced search costs for liquidity, fewer arbitrage-producing price discrepancies, fewer front-running opportunities (see Chapter 13), and potentially greater depth, the older concentrated markets suffered from many of the drawbacks that seem associated with monopoly, including higher trading costs. In the more fragmented markets, technology has improved (causal relationships go both ways), and market innovations are both more substantial and more quickly implemented. Thus, transactions are more quickly executed and their costs have diminished considerably.

2.4 THE DECLINE OF BRICK AND MORTAR

The natural monopoly power associated with the old exchange-based, open-outcry stock markets owes its early existence to a nearly complete absence of technology and communication systems. Investors were far too dispersed to conduct unaided securities transactions, and security traders required a central physical meeting site for competitive executions. This site or exchange enabled large networks of local broker offices to execute securities transactions on behalf of these widely dispersed investors. New York's Wall Street district provided an excellent venue for this purpose and the regional exchanges filled specific niches in the securities markets. As telecommunication technology advanced in the early decades of the 20th century offering market access to millions of investors, U. S. securities markets nearly self-destructed, only to be salvaged by regulatory authorities and intense efforts by exchanges to curb market abuse. Exchanges have since facilitated monitoring for regulatory compliance in much the same manner that they facilitate trading. Shortly after passage of the Securities Exchange Act of 1934, the SEC gained authority over the exchanges that, in turn, monitored and regulated their members.

Until the mid-1970s, brokerage firms maintained their monopoly power with a system of fixed transaction fees. All of their transactions were routed through the principal exchanges and off-floor markets. Exchanges had limited competition for securities listings. The primary sources of competition among securities brokers were in the arrays of services that they offered. After price controls were lifted in 1975, the market saw formation of a number of discount brokerage houses that unbundled nontransaction services such as research and were able to offer drastically reduced execution costs. The market segmented to an extent, with some firms offering only transaction executions and others offering the full array of brokerage house services such as client advice. However, the vast majority of transactions were still routed through the principal exchanges and off-floor markets.

More recent technological developments in telecommunications, wireless communications, the Internet, and globalization have imposed enormous competitive pressures on U.S. securities exchanges. With the electronic NASDAQ system beginning in 1971, competition from foreign fully automated exchanges trading U.S. securities (e.g., Eurex),

development of fully automated exchanges in the United States (e.g., the ISE), and ECNs, such as the Instinet, Island, and Posit systems that allowed direct trading between institutions, exchanges have been forced to accept and even innovate technological development to survive.

The natural monopoly enjoyed by traditional brick-and-mortar markets had been enhanced by a number of barriers to entry. Overhead outlays were required for large investments in office facilities, communications equipment, fixed exchange fees and memberships, broker training and licensing, and building client bases. Each of these overhead outlays was accompanied by significant time lags and increasing time required for recapturing initial investments (payback periods). In addition, traditional brokerage offices experienced significant returns to scale (see Stigler, 1961), making it almost impossible for smaller and newer firms to compete against larger better-established firms. Each of these barriers to entry into the securities brokerage business applies to the creation of overhead-intensive, brick-and-mortar securities exchanges and markets. For example, in the late 1990s, the Chicago Board of Trade and the New York Mercantile Exchange constructed new futures pits that cost $180 million and $228 million, respectively. However, recent electronic and other technological developments, particularly the Internet, substantially relieve these costly barriers to entry. First, most prospective clients already have Internet access and usage skills. Transaction costs are substantially reduced through use of equipment already installed in client homes. Whereas the typical U.S. customer stock transaction executed through a phone call has a variable cost of about $1, its cost is reduced to just $0.02 executed online. Similar disparities exist in open-outcry and electronic markets. For example, the open-outcry LIFFE charged about $1.50 per for long and short positions on the Bund futures contract in the late 1990s; its electronic German competitor DTB charged only $0.66 for the same contract. Open-outcry exchanges in the United States charged fees of approximately $1.50 per contract (see Cavaletti, 1997).

More significantly, overhead expenses for Internet brokers run approximately 1% of assets, as opposed to 2% to 3% for brick-and-mortar offices. Time required for Internet brokerage firm start-ups is substantially reduced relative to brick-and-mortar start-ups. In addition, the Internet and other technological advances have reduced economies of scale for brokerage firms. For example, as the Internet has largely replaced “cold calling,” the fixed costs of seeking and soliciting the business of small clients have dropped significantly. Furthermore, clients of electronic brokers trade far more frequently than those of full-service brokers, with some estimates ranging to 10 to 25 times as frequent (Varian, 1998). Hence, smaller individual investors are able to play larger roles in securities markets at lower costs. The increased participation of smaller investors, acting as day traders, enables them to replace specialists and market makers, at least to some extent, in the provision of liquidity to the market. In addition, the reduction of scale economies has increased competition among brokerage firms, in large part because so many of their services (e.g., advice, loans, and cash management) could be unbundled and commoditized through automation. Some of these services require very little initial capital outlay and no unique technology.

Removal of fixed brokerage commissions has substantially increased competition in the brokerage industry as reduced costs of service provision have softened barriers to entry. Brokerage commissions and fees had fallen from an average of $52.89 per trade in early

1996 to $15.67 in mid-1998. By 2000, a few online brokerage services had temporarily reduced their commissions to zero, and most still charge less than $15 per transaction. In fact, many stock brokers, including Robinhood Financial and Public.com, charge no commissions at all on stock trades (see end-of-chapter Exercise 4 for a discussion of this).[3] This particular scenario is interesting and quite controversial because it seems to be in part a direct result of *payment for order flow*, where brokerage firms and other institutions receive compensation from market makers (including high-frequency trading firms and internalizers) acting as wholesalers, electronic exchanges, market makers, and ATSs as payment for directing their order flow to those institutions or markets. However, this practice has grown more widespread, and in 2010, even the NYSE began making payments for order flow to providers of liquidity (traders whose orders that remove dealer inventory or orders from market backlogs). Barriers to entry based on ownership of physical facilities are disappearing, and existing firms are being forced to vary their product lines and merge into other institutions.

Combinations, alliances, and mergers between exchanges have also enabled previously smaller and less efficient markets to compete against larger ones. These combinations have enabled exchanges and markets to mutually benefit from one another's technological resources. Some of these combinations have been domestic such as OneChicago, LLC, an alliance created by the Chicago Board Options Exchange and Chicago Mercantile Exchange to trade equity futures contracts and the Pacific Exchange and the Archipelago Exchange merger. On the international scene, the New York Mercantile Exchange opened a satellite open-outcry trading floor in Dublin for trading in a Brent crude oil futures contract and mergers among exchanges in other countries such as those forming Euronext and Eurex (combining the Deutsche Börse AG/SWX Swiss Exchange). Each of these latter combinations offered improvements in investors' abilities to trade on a global basis.

Electronic versus Open Outcry

Which trading and brokerage platforms result in lower trading costs? For example, does open outcry result in superior trading performance relative to electronic execution? In an interesting study, Bakos et al. (2005), with $60,000 provided by the Salomon Brothers Center at New York University's Stern School, opened a series of accounts at various full-service, discount and electronic securities brokers. Their commissions for 100-share lots averaged $7.50 for electronic brokers and $47 for full-service voice brokers. They found that full-service brokers were more likely to route orders to the principal exchanges than electronic brokers and that such orders were more likely to be improved. However, for smaller orders, these price improvement advantages are more than offset by the higher brokerage commissions. Hence, specialists and market makers on exchanges were able to provide better order executions while brokers using electronic markets charged smaller commissions. It appeared that smaller investors fared better with discount electronic brokers while larger transactions resulted in better after-commission executions on the principal exchanges.

[3]Actually, as of the end of 2021, Public.com employed an optional compensation system of "tipping" for brokerage services as an alternative to required commissions.

Many of the studies of electronic versus open-outcry trading have used bid-offer spreads as a liquidity metric (e.g., Battalio, Greene, & Jennings, 1997), where wider spreads indicate lower liquidity. Shyy and Lee (1995) found that spreads appeared to be wider in electronic markets than in open-outcry markets; Pirrong (1996) found just the opposite. Pirrong argued that miscommunications and misunderstandings between trade participants reduce efficiency of open-outcry markets and that these issues are avoided in electronic markets. Regardless, time to execute trades is certainly reduced in electronic markets. In addition, screen-based trading has facilitated after-hours markets. Now, a number of exchanges are offering investors opportunities to trade after normal business hours.

The evidence concerning the provision of liquidity by open-outcry exchange markets relative to electronic markets is both ambiguous and mixed (e.g., Breedon & Holland, 1997; Pirrong, 1996). Some electronic systems have effectively eliminated driven floor-based, open-outcry markets. For example, Breedon and Holland describe how during 1997–1998, the computerized Eurex drew practically 100% of the trading in German Bund futures from the open outcry LIFFE, which had held a 70% market share only months earlier. By 2000, LIFFE had abandoned open outcry entirely for LIFFE CONNECT, a fully automated system. However, LIFFE did retain the bund options contracts, probably because of the more complicated strategies associated with them. In 1998, MATIF operated its open-outcry and electronic systems simultaneously. Within two weeks, the computerized system had taken all volume from the open-outcry markets that had to be closed. The Hong Kong Futures Exchange and the Sydney Futures Exchange both abandoned open outcry during this same period. Frino, Hill, Jarnecic, and Aitken (2004) found that spreads narrowed on the Sydney and Hong Kong Exchanges, whereas spreads widened on the LIFFE. In addition, according to Frino et al., bid-ask spreads on all three exchanges appear to widen in response to price volatility at a faster rate under electronic trading than with open-outcry trading, suggesting that the specialist-based system may offer better price continuity in periods of uncertainty.

One of the difficulties of screen-based trading is the inability to efficiently disseminate significant amounts of information concerning trades. While it is easy enough for screens to display bid and offer prices, open outcry participants are able to more easily verbally communicate order types and combinations as well as other more complicated trade details. Such matters grow in importance when trade sizes are larger or when, for example, an options trader is attempting to leg into a spread (execute spread transactions one at a time) or other position. While screen-based trading provides for a greater level of anonymity, many traders prefer a market where their counterparties can be identified. In addition, Sarkar and Tozzi (1998) argue that open-outcry exchanges provide for more liquidity in more active markets while newer, less active issues are less likely to be found in open-outcry markets.

One of the more controversial aspects of electronic brokerage transactions is that customers do not normally exercise their right to direct trade order routings and that orders are not routed in their entireties to the markets with the best prices. Brokers do have a responsibility to route orders to the market with the best execution (or price, depending on the particularly regulatory regime), a rather ambiguous and difficult concept to implement, and one that few retail traders understand. Many brokers have agreements with order wholesalers or market makers such as KCG Holdings, Inc. and Citadel Securities, LLC.,

which in turn, route orders so as to internalize them (execute against their own inventory) to capture larger spreads. Other orders might result in compensation by exchanges and markets to route transactions to them. These exchanges and market makers pay wholesalers for order flow that result in profits to wholesalers, creating conflicts of interest in dealings with clients. This practice, "payment for order flow," was pioneered by the infamous Ponzi schemer Bernard Madoff.[4] Thus, a retail brokerage firm will receive payments from a wholesaler that receives payments from an exchange or other market for routing orders to that market. For example, in 1999, the Knight/Trimark Group was paid $138.7 million for order flow, with over 10% of this sum received from Ameritrade, a large electronic brokerage firm (Bakos et al., 2005). TD Ameritrade recorded revenue of approximately $1.15 billion in 2020 from payments for order flow. Specialists and many market makers on the major exchanges resented this controversial practice of order flow payment and many discount retail brokerage clients seem unaware of it. But, as we discussed earlier (Section 2.4), in early 2010 even the NYSE began making payments for order flow to providers of liquidity. While the conflicts of this practice are obvious, payment for order flow does subsidize brokerage operations, which in turn, might reduce client trading costs. In addition, the practice might enhance liquidity. However, Rosov (2016) argues that resultant conflicts of interest might more than offset the reduced trading costs.

2.5 CROSSING NETWORKS AND UPSTAIRS MARKETS

One of the most vexing problems faced by institutional traders is the price pressure, called *slippage*, associated with large transactions. For example, placing a large sell order will force prices down against the seller. In fact, merely attempting to execute a large sell order reveals negative information to the market. To avoid adverse price effects of orders, traders often use *crossing networks*, which are alternative trading systems that match buyers and sellers with respect to agreed-upon quantities. Crossing networks do not publicly display quotations, thereby enabling participants some degree of anonymity and insulating them from slippage. Trades are priced by reference to prices obtained from other markets rather than by the more traditional auction systems.[5] Traders using crossing networks announce their order sizes. The crossing network then matches buy and sell orders at prices obtained from more traditional auction markets such as the NYSE. This crossing procedure enables institutional traders to execute without exposing their order flow to competitors. The crossing network, by having prices determined elsewhere, also prevents the price impact or pressure typically associated with auctions of large blocks of shares. Because crossing networks do not reveal prices or client identities, and they represent nondisplayed liquidity, they are sometimes referred to as *dark pools of liquidity*. In sum, these dark pools of liquidity accounted for 37% of U.S. equity volume in August 2014 (Baldauf & Mollner, 2020).

[4]We will discuss payment for order flow and Bernard Madoff separately in Chapter 13.

[5]The Securities and Exchange Commission (1998) defines crossing networks as "systems that allow participants to enter unpriced orders to buy and sell securities. Orders are crossed at a specified time at a price derived from another market" (i.e., a continuous market such as the NYSE).

As initially established, crossing networks provided ideal marketplaces for large passive investors such as index funds to transact for large blocks of securities, at prices based on published auction market bid-ask quotations, while avoiding slippage problems. However, in more recent years, several networks have reported smaller average transactions sizes as traders find a wider variety of uses for these trading venues. Several of these crossing networks have accommodated traders by providing electronic executions, many based on algorithms to determine order execution sizes and prices.

The NASDAQ Crossing Network provides traders with opportunities to match orders at known benchmarks (e.g., the mid-point of the spread) or closing or other prices several times a day. For example, trades can be executed after regular market hours, providing liquidity outside normal trading sessions. The match price might be the volume weighted average price (VWAP) of all trades during the day (see Section 10.3 for details on VWAP). ITG's Posit systems allow traders to enter order quantities that are matched during listing market trading sessions with execution prices at the midpoint of best bid and offer currently prevailing in the listing market.

We discussed *Liquidnet* in Section 1.3, which is a negotiating venue that matches institutional buyers and sellers of large blocks of equity securities, enabling them to directly bargain and trade confidentially with one another. Liquidnet provides an electronic format for this negotiation process. Traders use the system by entering symbols for securities that they wish to buy or sell. If Liquidnet finds a potential match, it notifies the counterparties, who bargain anonymously with each other in a virtual meeting room using the Liquidnet system. This facility for bargaining distinguishes Liquidnet from the other crossing systems discussed here. Thus, trade counterparties negotiate directly with each other the terms of their trade. Counterparties will know prices from other markets, leaving relatively little room for significant pricing disagreements. If they come to terms, they report the transaction to Liquidnet, which arranges the transfer and takes a commission on each share.

The New York Stock Exchange also conducts several after-hours crossing markets. First, the NYSE offers Crossing Session I from 4:15 to 5:00 pm after the NYSE closes regular floor trading. Crossing Session I, when in effect, enables traders to match single stock orders at closing price. In addition, the NYSE offers Crossing Session II from 4:00 to 6:30 pm to enable program trades for 15 or more securities, regardless of their value. Furthermore, the NYSE MatchPoint system was an electronic facility that matched aggregated orders at predetermined fixed times with prices that were derived from ongoing floor trading.

Internalization

Internalization occurs when brokers execute their own client buy orders against their own client sell orders, representing both sides of a trade and without routing them to central markets. Internalization allows brokers to execute transactions more easily and at a lower cost. Of course, internalization does present some obvious potential problems. First, internalization may inhibit the broker's ability to properly represent the client as the client's agent, particularly when the broker is also representing a client on the other side of the transaction or acting as a dealer. Second, internalization results in fewer transactions being executed in the central market, which increases market fragmentation and

reduces transparency and potential price competition. These problems can lead to reduced liquidity and increased price volatility in the central market. Internalization can also lead to violations of price and time priority. Perhaps worse for the broker's clients, the transaction may be more susceptible to manipulation or may not be executed at the best possible prices, consistent with the point above concerning the broker's ability to represent the client.

Internalization of customer orders is not possible for options markets transactions when options transactions must be executed on options exchanges as per rules of each options exchange. However, some brokerage firms will internalize or submit their own orders simultaneously with their customer orders, such that there is a reasonable possibility that they will execute against each other. On the other hand, the options exchange trading crowd might "break up" one or both sides of the orders, perhaps at better prices, so that the brokerage firm only participates in the executions of parts of one or both orders.

2.6 FIXED-INCOME SECURITIES AND MONEY MARKETS

In terms of dollar volume, the U.S. markets for debt instruments are larger than for any other type of security. Debt markets, including markets for mortgages, are more than twice as large as stock markets. Debt securities, sold for the purpose of raising capital, are IOUs issued by a variety of types of organizations, including federal, state, and local governments and their agencies, as well as by corporations and other institutions. A debt security represents a claim on the issuer for a fixed series of future payments. For example, a debt security might specify for its owner to receive a stated series of interest payments until the instrument matures. The instrument might also provide for principal repayment when the instrument matures. These securities are issued in *primary markets* and then traded in *secondary markets*, just as are other financial instruments. Debt securities will normally specify terms of payment, including amounts and dates, collateral and priority (who is to be paid first in the event of issuer financial distress). Among the various types of debt securities are bonds, notes, mortgages, and Treasury instruments. Many of the fixed income securities with shorter terms to maturity are considered to be money market instruments.

U.S. Treasury Securities and Markets

The U.S. Treasury is the largest issuer of debt securities in the world. In 2020, the U.S. Treasury (technically, the Fed) auctioned $19.683 trillion in Treasury instruments through 503 public auctions using the Treasury Automated Auction Processing System (TAAPS). The federal government raises (borrows) money through the sale of U.S. Treasury issues including Treasury bills (T-Bills), Treasury notes, Treasury bonds, Treasury inflation-protected Securities (TIPS), and Floating Rate Notes (FRNs). By purchasing Treasury issues, an investor is loaning the government money. Notwithstanding the 2011 debt ceiling crisis, and several threats of repeat crises, the U.S. federal government has proven to be an extremely reliable debtor (at least it has made good on all of its Treasury obligations

with nearly 100% consistency). Treasury issues are fully backed by the full faith and credit of the U.S. government, which has substantial resources due to its ability to tax citizens, create money and borrow more. Thus, these securities are generally expected to have the lowest default risk and should generally be safer than the safest of corporate bonds or short-term notes.

Aside from occasionally issued cash management bills (CMBs) with maturities as short as 1 day, the Treasury obligations with the shortest terms to maturity are *Treasury bills*. They typically mature in less than one year (4, 13, 26 or 52 weeks). These issues are sold as pure discount debt securities, meaning that their purchasers receive no explicit interest payments. Instead, investors purchase them at a discount from their maturity or face values. Such pure discount instruments are also known as zero-coupon issues.

One variation on the T-Bill issue is a so-called *STRIP*, issued through the U.S. Treasury's *Separate Trading of Registered Interest and Principal Securities* (STRIPS) program beginning in 1985. While by no means identical to T-Bills, STRIPS are portfolios of single-payment instruments sold by the Treasury in blocks with varying maturities. For example, a single STRIP maturing at the end of five years from now would provide for a fixed payment at that time. Individual STRIPS sold in blocks can be "stripped" from the block and sold in secondary markets.

In addition to the short-term pure discount instrument issues discussed above, the Treasury also offers a number of longer-term coupon issues. For example, *Treasury notes* (T-Notes) have maturities ranging from 1 to 10 years (typically, 2, 3, 5, 7 and 10 years) and make semi-annual interest payments. Similarly, *Treasury bonds* (T-Bonds) typically range in maturity from 10 to 30 years and make semi-annual interest payments. Bonds issued prior to 1985 were frequently callable, meaning that the Treasury maintained an option to repurchase them from investors at a stated price. The last such call occurred in 2009. More than half the marketable Treasury debt outstanding as of 2000 was in the form of notes, while bills and bonds each represented about 20%.

Treasury inflation-protected securities (TIPs) are currently offered with 5-, 10-, and 30-year maturities, and have coupon and principal amounts adjusted by the Consumer Price Index. Floating rate notes are currently offered with 2-year maturities and have coupon rates indexed to the most recent 13-week Treasury bill offering. The U.S. Treasury also offers nonmarketable issues such as *Series EE U.S. Savings Bonds*, *Series I U.S. Savings Bonds*, and *Series H U.S. Savings Bonds.* These savings bonds are normally issued only to individuals and cannot be traded among investors. Such issues are often subject to certain restrictions (such as a $5000 maximum level of Series EE purchases per year). These bonds can be purchased through most banks and savings institutions, and many businesses maintain plans through which their employees can purchase savings bonds through payroll deduction programs.

Agency and Government-Sponsored Enterprise Issues

The U.S. federal government has created and sponsored a number of institutions known as *agencies* and *government-sponsored enterprises*. These institutions enable the government to make funds available for a number of policy-related functions. Among the oldest of these government-sponsored enterprises is the *Federal National Mortgage Association*

(FNMA or Fannie Mae), created in 1938 by the Federal Housing Administration (FHA) to expand the flow of money to housing markets during the Great Depression. This institution was intended to spur investment into real estate, improve employment during the depression, and to help enable people to purchase homes. The primary functions of Fannie Mae were to purchase, hold, and sell FHA-insured (and, after World War II, Veterans Administration−insured) mortgage loans originated by private lenders. Created with a congressional charter, and then privatized in 1968, FNMA was, until September 2008, one of the largest privately owned corporations in the United States with shares traded on the New York Stock Exchange. However, the institution was delisted and taken over by the federal government during the financial crisis of 2008. The firm received a bail-out package from the federal government to support it until 2012, and remains owned by the federal government.

FNMA facilitates capital acquisition in the mortgage industry through the creation of *mortgage-backed securities*, instruments denoting ownership in these pools of mortgages. Mortgage-backed securities are created by their sponsors who purchase residential mortgages from banks and thrift institutions and then create debt and other securities backed by pools of these mortgages. In effect, FNMA purchases the mortgage obligations held by banks and thrifts, repackages them as debt security portfolios, insures them, and re-sells them to the general public. The funds raised by the sale of these securities are then used to purchase additional mortgages from banks, increasing capital available to the mortgage and housing markets. These portfolios of mortgage-backed securities are also called *pass-through securities* because the interest and principal obligations associated with the mortgages are passed through to the security holders and enable banks with relatively short liability terms to remove 30-year mortgage assets from their balance sheets.

Typical issues of government-sponsored enterprises (GSEs) such as those of FNMA do not have the full faith and credit backing of the federal government as do the bulk of agency issues such as those of GNMA. Separating and repackaging the credit risk implications of FNMA issues through the process of synthetic securitization enhances the market potential of GSE issues. Credit risk derivatives such as synthetic collateralized debt obligations (CDOs) and credit default swaps (CDSs) created through this process are marketed separately from the mortgages with which they were initially created.

The *Government National Mortgage Association* (GNMA or Ginnie Mae), established by Congress in 1968 as a "spin-off" of FNMA, is a corporation owned by the U.S. federal government. GNMA guarantees mortgage-backed securities for the FHA (Federal Housing Administration), Veterans Administration (VA), and other agency mortgage issues. Many mortgages issued by these agencies are targeted towards particular groups such as lower-income or veteran families, and experience relatively high default rates. The mortgages issued by agencies, such as GNMA, unlike government-sponsored enterprises such as FNMA, are *full faith and credit instruments*, meaning that the U.S. federal government backs them with its full ability to tax, borrow, and create money.

The *Federal Home Loan Mortgage Corporation* (FHLMC or Freddie Mac), a stockholder corporation created in 1970 by the federal government, also creates, insures, services, and sells pass-through securities related to single family and multifamily residential mortgages. Freddie Mac's creation was essentially to provide competition in the secondary mortgage market to the virtual monopoly granted to Fannie Mae, thanks to federal

protections enjoyed by Fannie Mae. Freddie Mac and Fannie Mae were intended to compete in a "separate but equal" regulatory framework. Freddie Mac was also placed in conservatorship by the federal government during the financial crisis of 2008–2009. Freddie Mac is currently a government-owned corporation.

The *Student Loan Marketing Association* (SLMA or Sallie Mae) was established in 1972 to create, insure, and sell pass-through securities related to student loans. While originally created as a government agency in 1972, the company began a privatization process in 1997 that was completed in 2004. The institution is now a publicly traded firm known as *SLM Corporation*, and originates, services, and collects private student loans. In 2014, the firm spun off its federal student loan servicing businesses into a separately traded company known as Navient Corporation.

Municipal Securities and Markets

The municipal securities markets owe much of their success to the U.S. federal taxation code that permits investors in municipal instruments to omit from their taxable income any interest payments received on these issues. Thus, as of 2021, interest received on municipal bonds normally need not be declared as part of personal income subject to federal income taxation. This feature makes municipal bonds more attractive to investors, particularly those in higher income tax brackets, enabling state and municipal issuers to offer these bonds at reduced interest rates. However, such municipal bonds are more difficult to market to tax exempt institutions, pension funds, individuals in low income tax brackets and mutual funds because they do not receive enough benefit from the income tax exemptions to offset their low interest rates.

Several types of municipal issues are offered by state and local governments. The first, so-called *general obligation bonds*, are full-faith and credit bonds. This means that the issuer backs the bonds to the fullest extent possible, given its assets and other obligations. However, municipal issuers usually lack the unlimited taxing and money creation abilities of the federal government. Thus, default risk analysis and ratings are important. *Limited obligation bonds* provide for the issue to be backed only by specific resources or assets. For example, a revenue bond may be backed only by the cash flows generated by a specific asset such as a toll bridge.

Most municipal issues are rated for default risk by private rating agencies such as *Fitch*. Many municipal issues are insured by private insurance institutions. This insurance is intended to mitigate the default risk associated with the issue, and make them more attractive to investors. This reduced risk enables issuers to offer bonds with reduced interest rates to the public. Among the larger insurers of municipal bonds are the AMBAC Assurance Corporation, the National Public Finance Guaranty Corp., the Financial Guaranty Insurance Company (FGIC), MBIA Insurance Corporation, and the Assured Guaranty Municipal Corp.

While most municipal issues are income tax exempt at the federal level, and, under certain circumstances at state and local levels as well, not all are tax exempt, and some municipal issues qualify for other types of benefits. For example, the Build America Bonds (BABs) program, created to help revive the American building industry during

the 2009–2010 recession, make taxable interest payments to bondholders, but allow for issuers to receive either 35% federal subsidies on interest payments that they make or, instead, offer special tax credits for the bondholder. As with typical municipal bonds, these federal programs reduce the net cost of financing with these municipal bonds. These long-term bonds are often more marketable than their tax-exempt counterparts to tax exempt institutions, pension funds, and mutual funds because their interest rates are higher. Somewhat similar taxable municipal issues include tax credit bonds, which include qualified school construction bonds, qualified zone academy bonds, clean renewable energy bonds, and qualified energy conservation bonds. These bonds offer similar subsidy or tax considerations.

Money Market Issues of Financial Institutions

Nongovernment financial institutions are obviously significant issuers and borrowers in the primary markets for debt instruments. For example, the *federal funds* markets allow banks and other depository institutions to borrow from one another to meet Federal Reserve requirements. Essentially, this market provides that excess reserves of one bank may be loaned to other banks for satisfaction of U.S. central bank reserve requirements. The rate at which these loans are extended is referred to as the *federal funds rate*.

Normally, bank accounts are not regarded as marketable securities. One exception to this is the *negotiable certificates of deposit* (a type of jumbo CD). This is a tradable depository institution CD account with a denomination or balance exceeding $100,000. The amounts by which jumbo CDs exceed $250,000 are not covered by FDIC insurance. Negotiable CDs are often purchased by wealthy individual investors, financial institutions, corporate treasuries, and by *money market mutual funds*, which are created by banks and investment institutions for the purpose of pooling together depositor or investor funds.

Banker's acceptances are originated when a bank accepts responsibility for paying a client's loan or assuming some other financial responsibility on behalf of a client. Because the bank is likely to be regarded as a good credit risk, these acceptances are usually easily marketable as securities.

Repurchase agreements (repos) are issued by financial institutions (usually securities firms and banks) acknowledging the sale of assets (typically treasuries and fixed-income instruments) and a subsequent agreement to repurchase at a higher price in the near term. This agreement is essentially the same as a collateralized short-term loan; the securities (collateral) reduce credit risk, and the ability of the buyer to be the seller in subsequent repo agreement reduces liquidity risk. Both reductions in risk serve to reduce the cost of the loan to the original repo seller. The repo rate is determined by the price at termination relative to the price at the initiation of the agreement, then adjusted for time. The counterparty institution buying the securities with the agreement to resell them is said to be taking a *reverse repo*.

Corporate Bonds and Markets

Corporations are also important issuers of debt securities. Large, well-known, creditworthy firms, including financial institutions, needing to borrow for a short period of time

may issue large denomination short-term notes frequently referred to as *commercial paper*. Well-developed markets exist for these short-term promissory notes, though secondary markets tend to be inactive. Firms requiring funds for longer periods of time may issue *corporate bonds*. Many issues are rated by major credit rating agencies, as we will discuss later. These longer-term instruments are often issued with a variety of features, including callability, convertibility, sinking fund provisions, and so on. There are many different types of corporate bonds. The terms of the bond will be specified in a contract known as a *bond indenture*. In addition, firms may make bank commercial loans, although secondary markets for bank loans tend to be limited in size and scope.

Callable bonds can be *called* by the issuing institution at its option. This means that the issuing institution has the right to pay off the callable bond before its maturity date. The callable bond typically has a *call date* associated with it as well as a *call price*. The call date is the first date (and perhaps only date) that the bond can be repurchased by the issuing institution. The call price is normally set higher than the bond's par value and represents the price that the issuing institution agrees to pay the bond owners. Because the issuing institution retains the option to force early retirement of callable debt, the call provision can be expected to reduce the market value of the callable bond relative to otherwise comparable noncallable bonds.

Convertible bonds can be convertible by bondholders into equity or other securities. This normally means that the convertible bondholder has the right to exchange the convertible bond for a specified number of shares of common stock or some other security. The convertibility provision of such a bond enhances its value relative to otherwise comparable nonconvertible bonds.

Debentures are not backed by collateral. However, they and other bonds can have a *sinking fund* provision to provide for additional repayment assurance for bondholders. One type of sinking fund provision has the issuing institution placing specified sums of money into a fund at specified dates that will be accumulated over time to ensure full satisfaction of the firm's obligation to bondholders. Alternatively, sinking funds will be used to retire associated debt early. The debt retirement can be effected either in the open market at prevailing market prices or called back by the issuing firm. The call can occur at the call price if specified in the bond indenture or, if so provided by the indenture, called at random by the issuing firm. *Serial bonds* are issued in series with staggered maturity dates.

Many more innovative bonds have been offered in the market. *Floating rate* bonds have coupon rates that rise and fall with market interest rates; *reverse floaters* have coupon rates that move in the opposite direction of market interest rates. *Indexed bonds* have coupon rates that are tied to the price level of a particular commodity like oil or some other value such as the inflation rate. *Catastrophe bonds* make payments that depend on whether some disaster occurs, such as an earthquake in California or a hurricane in Florida. These catastrophe bonds provide a sort of insurance for issuers against the occurrence of the disaster. In some respects, purchasers of these bonds are providing insurance to the issuers.

The predecessor to the NYSE was established in 1792 as a venue for trading bonds. Since then, corporate and later municipal and Treasury bonds have been traded on organized exchanges and in the OTC markets, with the relative importance of these markets changing over time. While over 1000 bonds trade on "NYSE Bonds," the largest centralized corporate bond market in the United States, the majority of trades for corporate bonds

occur in the OTC markets. FINRA TRACE (Trade Reporting and Compliance Engine) provides corporate bond trading data to the public, as all FINRA member broker/dealers are required to report OTC corporate bond transactions data to TRACE.

Credit Ratings and Credit Agencies

Most publicly traded corporate bonds are rated by well-known agencies with respect to anticipated default risk. Corporations pay institutions such as Standard & Poor's and Moody's to rate the risk of their issues, although a few credit agencies, including Egan-Jones, have received revenues from investor subscribers. Both of these credit rating agencies are registered with the SEC as NRSROs, Nationally Recognized Statistical Rating Organizations, which allow them to rate creditworthiness for purposes on regulatory suitability to be purchased by banks. Other U.S. NRSROs include Fitch, A.M. Best, HR Ratings, Japan Credit Rating Agency, LACE Financial Corp., Rating and Investment Inc., Morningstar Credit Ratings LLC, and DBRS Ltd. Bonds without ratings assigned by these agencies can be very difficult to sell; in fact, many institutions face regulatory restrictions on purchasing bonds that are either unrated by NRSROs or have ratings below a given level. These constraints provide issuing companies with incentives to seek ratings from NRSROs. In past years, particularly before the financial crisis of 2007–2009, reputations enjoyed by these rating agencies had generally been quite solid.[6] In fact, as indicated above, numerous international and federal government regulatory agencies including the SEC and the Fed sanction NRSROs, and constrain or limit banks and other institutions' investment in low-rated instruments. For example, Basel II guidelines allow for "external credit assessment institutions" ratings for assessing bank risk. The SEC allows investment banks and broker-dealers to use credit ratings from NRSROs listed above for similar purposes. In effect, rating agencies can serve as for-profit regulatory authorities. Federal regulatory agencies such as the FDIC can prohibit financial institutions from buying risky bonds, while credit agencies determine which bonds are risky. For example, federally chartered banks are often restricted from purchasing bonds with ratings below investment grade (BBB or higher are investment grade). Standard & Poor's and Moody's use the rating schemes depicted in Table 2.5, though the actual rating categories can be more finely graded with the addition of a plus or minus sign and/or with a numerical value. Other rating agencies use similar schemes.

While credit rating agencies have enjoyed many decades of strong reputation and success, many have been rounded criticized for their failures leading to the 2006–2008 U.S. housing crisis and the 2010/2011 European Union (EU) debt crisis. Many critics have argued that these agencies are often too slow in responding to dramatic shifts in market conditions and that traders are often much faster at responding to such shifts. For example, Enron bonds were rated triple-A by both major agencies until only a few days before the company filed for its 2001 bankruptcy protection. The agencies also seemed to miss troubles brewing at WorldCom,

[6]Some credit rating agency reputations were marred during and after the 2007–2009 financial crisis by conflicts of interest and their apparent lack of response to the increased risk of rated issues. The problems were particularly acute with respect to mortgage-related instruments and instruments with derivative security characteristics. These conflicts and slow responses are discussed shortly here and in Chapter 13.

TABLE 2.5 Standard & Poor's and Moody's Corporate Bond Ratings

Description	Standard & Poor's	Moody's
Least likely to default	AAA	Aaa
High quality	AA	Aa
Medium-grade investment quality	A	A
Low-grade investment quality	BBB	Baa
High-grade speculative quality	BB	Ba
Speculative	B	B
Lower-grade speculative	CCC	Caa
Highly speculative	CC	Ca
Likely bankruptcy	C	C
Already in default	D	D

Finer rating gradations noted with plus, minus and numerical values are omitted from this table.

Parmalat, and Delphi until just before their bankruptcy filings, as well as conditions leading to the 2006–2008 subprime crises, particularly with respect to mortgage-backed securities. In fact, several credit rating agencies maintain investment grade ratings for Lehman Brothers as late as the morning of its September 2008 bankruptcy filing. Nevertheless, bond rating agencies do play a crucial role in the evaluation of default risk.

Bonds rated BBB (or Baa) and higher are typically referred to as investment-grade bonds, while bonds below this level are considered speculative grade. Speculative-grade bonds are often called *junk bonds*. There is significant evidence that these bond ratings are highly correlated with incidence of default, suggesting that these agencies are least reasonably useful in forecasting default and indicating default risk. Furthermore, it is fairly unusual for ratings provided by these agencies to differ by more than one grade. Bond markets seem to agree with these statistical findings, pricing bonds such that their yields are strongly inversely correlated with bond ratings.

Credit rating agencies rate a variety of corporate, municipal, and financial institution issues. For example, S&P, Moody's, Fitch, and DBRS rate commercial paper. Moody's uses the lettering system P1, P2, and P3 to indicate superior, satisfactory and adequate ratings. S&P follows a similar system with A1, A2, A3, B, C, and D ratings, Fitch with F1 to F5 and DBRS with R1, R2, and R3.

Bond rating agencies make extensive use of financial statement and ratio analysis to compute their ratings. Such analyses are frequently supplemented by statistical techniques such as *multidiscriminate analysis*, *probit*, and *logit* modeling. While their ratings in the past have been almost universally accepted in financial markets, even to the extent that numerous SEC, Federal Reserve, and other regulations have been based on these ratings, there have been problems with the ratings:

1. Many rating agencies are paid by the firms that issue debt instruments, creating potential conflicts of interest.

2. Rating services have provided consulting services to issuing firms, advising them on strategies to improve their ratings. However, such services have been restricted by the Credit Rating Reform Act of 2006, Dodd-Frank and other legislation (see Chapter 5).
3. Credit agencies have been accused of strongarming clients and potential clients. For example, Moody's published an "unsolicited" rating of Hannover Re, a German re-insurance firm. It sent a letter to Hannover indicating that "it looked forward to the day Hannover would be willing to pay" (Klein, 2004). Hannover management refused payment, perhaps based on its paid use of two competing ratings agencies for its debt. Moody's continued to rate Hannover debt, which continued refusing to make payments. Over time, Moody's cut Hannover ratings to junk status, despite high ratings given by other agencies.
4. As mentioned above, rating agencies have missed major bond crises and other scenarios when bonds should have been downgraded.
5. Numerous studies, including Xia (2014), have suggested that credit agencies lag the market in re-evaluating credit risk and updating ratings. In fact, since debt issuers often pay for credit ratings, credit rating agencies might even have an incentive to delay credit rating downgrades to avoid displeasing revenue sources.

NRSROs are required to provide to the public performance statistics for the issues that they rate. These performance statistics are normally available on agency websites, and typically include default rates and transition (rating up- and down-grades) for their rated issues.

Eurocurrency Instruments and Markets

Eurodollars are freely convertible dollar-denominated time deposits outside the United States.[7] The banks may be non-U.S. banks, overseas branches of U.S. banks, or international banking facilities (not subject to reserve requirements). Eurodollar markets began after World War II when practically all currencies other than the U.S. dollar were perceived as unstable. Thus, most foreign trade between countries was denominated in U.S. dollars. However, the Soviet Union and Eastern European institutions were concerned that their dollars held in U.S. banks might be attached by U.S. residents in litigation with these countries. Thus, they dealt not with actual U.S. dollars, but merely denominated their debits and credits with dollars. Monies owed to them were simply offset by monies that they owed. In a sense, they dealt with "fake" eurodollars, but since their trading partners did also, and their accounts tended to "zero out" over time, this did not create significant problems. Their eurodollars were left in Western European banks. During the 1960s and 1970s, these markets thrived due to regulations imposed by the U.S. government such as Regulation Q (interest ceilings), Regulation M (reserve requirements), the Interest Equalization Tax imposed beginning in 1963 to tax interest payments on foreign debt sold in the United States, and restrictions placed on the use of domestic dollars outside the United States. More generally, eurocurrencies are loans or deposits denominated in currencies other than that of the country where the loan or deposit is created. Roughly 65% of eurocurrency loans are denominated in dollars.

[7]It is important to note that eurodollars and eurocurrencies are not euros, the currency used in the EU.

Eurocredits (e.g., eurodollar credits) are bank loans denominated in currencies other than that of the country where the loan is extended. They are attractive due to very low interest rate spreads that are possible due to the large size of the loans and the lack of reserves, FDIC, and other requirements directly or indirectly affecting domestic loans and deposits. Their rates have traditionally been tied to the LIBOR (London Interbank Offered Rate, based on a survey panel of interbank loans) and various U.S. rates that will be discussed later. However, there has been a widespread migration in contracts from use of LIBOR as a benchmark towards the *Secured Overnight Financing Rate* (*SOFR*), which is based on calculations drawn from the far more extensive U.S. Treasury repo markets. The U.S. Federal Reserve has announced that the LIBOR should be phased out and replaced by June 30, 2023.

Loan terms for Eurocredit instruments are usually less than five years, typically for six months. *Eurocommercial paper* is the term given to short-term (usually less than six months, often just a few days) notes issued by large, particularly "credit worthy" institutions. Most commercial paper is not underwritten. The notes are generally very liquid (much more so than syndicated loans) and most are denominated in dollars. They are usually pure discount instruments. EMTNs are interest-bearing instruments usually issued in installments. Most are not underwritten. Eurobonds are generally underwritten bearer bonds denominated in currencies other than that of the country where the loan is extended.

2.7 MARKETS AROUND THE WORLD

As discussed above, the oldest continuously operating stock exchange is Euronext Amsterdam Bourse, which first traded stock of the East Indies Trading Company in 1602.[8] The NYSE started trading in the streets; the Paris Exchange, now known as Euronext Paris, organized under Napoleon's regime, started on a bridge (Pont au Change); and the London stock market started in a tavern. Traditional brick-and-mortar exchanges since have been any one of three basic types. The first, the *public bourse*, is essentially a public institution, established by and operating under the authority of a government. However, most public bourses, including the Paris Bourse in 1990, have reorganized to become private bourses. The *private bourse*, such as the NYSE, is a private institution established by independent investors and institutions. The NYSE has evolved into a publicly traded stock institution, although it is still a private bourse in this context. These exchanges are still subject to governmental regulation. Finally, *bankers' bourses* exist where the primary dealers of securities are banks, and where commercial banking functions were not distinct from brokerage functions as was the case in the United States. In Germany, bankers' bourses have served banks who were also the primary stockbrokers.

As of 2021, the largest stock markets in terms of monetary volume are in New York (NYSE and NASDAQ), China, Europe, and the U.K. As of 2021, the NYSE listed equities exceeding $26 trillion in market value (see Table 2.6), with Nasdaq gaining ground in recent years.[9] The United States and United Kingdom (UK) have long histories of funding business with publicly traded equity securities, whereas commercial bank funding has

[8]Some material in this section was adapted from Solnick (1991).

[9]Figures in this section and Table 2.6 are adapted from reports by the World Federation of Exchanges.

TABLE 2.6 Major Equity Exchanges of the World

	Exchange	US$ Trillion
1	NYSE (US)	26.64
2	NASDAQ (US)	23.46
3	Shanghai Stock Exchange (China)	7.63
4	Euronext (Europe)	7.33
5	Japan Exchange Group	6.79
6	Hong Kong Exchange	6.13
7	Shenzhen Stock Exchange (China)	5.74
8	LSE Group (UK)	4.05
9	Bombay Stock Exchange (India)	3.96
10	National Stock Exchange (India)	3.77

Market Capitalization in USD Trillions, 2020/2021.
Source: World Federation of Exchanges.

played larger roles in Japan and Germany. This is one of the major reasons as to why the London stock markets are larger than those elsewhere in Europe, despite the relative small size of the UK economy relative to that of Europe.

Since 1974, China, India, and the Pacific region have produced the largest growth rates in equity markets as a percentage of world equity markets. The combined proportions of these markets as a percentage of world markets have grown from 16% to 32% in 2010. Japan, China and India have been the most significant contributors to this figure. European equity markets as a percentage of the world total have increased from 22% to 24%. U.S. equity markets have remained fairly constant, declining from 57% to 55% of the world total.[10] Over this 43-year period, the world equity market has grown from less than $900 billion in 1974 to over $69 trillion as of 2020.

Emerging markets are those that have some but not all of the characteristics of developed markets. While the term's definition varies, and some economists even argue about whether the term still is useful, there is still some agreement as to where the term might apply. The BRIC countries, Brazil, Russia, India, and China, are considered to be the largest of the emerging economies. Notice in Table 2.6 that China and India are represented in the world's largest equity markets. Other emerging markets might include South Korea (though, it is arguably a developed country), Mexico, Indonesia, Turkey, Saudi Arabia, Iran, and South Africa. The Economist and Morgan Stanley Capital International (MSCI) compile well-known lists and other information on emerging markets.

Reports from ING Baring Securities and the U.S. Census Bureau indicated that U.S. ownership of foreign stocks has increased from $41 billion in 1985 to $3.977 trillion in 2010.[11] This

[10]These percentage figures are estimates, drawn from somewhat different periods and sum to more than 100%.

[11]See the *Wall Street Journal*, May 28, 1996, p. R23, and U.S. Census Bureau, Table 1203, "U.S. Holdings of Foreign Stocks and Bonds by Country," 2011 Statistical Abstract.

increased investment has been the product of a number of factors, including improved investing technology, reduced constraints on capital mobility, a greater appreciation for benefits of diversification, and a belief in the potential for higher returns from foreign investment. Historically, European and Japanese investors have maintained higher levels of international investment, but U.S. equity portfolios have significantly increased foreign holdings.

2.8 CURRENCY EXCHANGE AND MARKETS

For transactions to be executed between countries with different currencies, some type of foreign exchange (also called currency exchange, forex, or FX) must take place. Foreign exchange is simply the trading of one currency for another. For example, an American tourist in France can exchange dollars from her bank account for euros at an automated teller machine (ATM) in Paris.

Markets for foreign exchange are crucial for trade and the economies of most countries. The total monetary volume of foreign exchange markets is huge. The Bank for International Settlements, a major regulator and facilitator in markets for foreign exchange, estimated that the market for foreign exchange was approximately $1.2 trillion per day in 1996, growing to over $6.6 trillion daily by April 2019, making the exchange market the largest financial market in the world, larger than all world equity and bond markets combined.[12] Approximately $2.4 trillion of this volume involved spot contracts; the remainder involved forward, swap, and other derivative contracts. Almost two thirds of transactions involved counterparties trading from different countries.

Why exchange currencies? Obviously, currency exchange is needed in many international transactions for consumption and investment. For example, a Canadian automobile leasing firm purchasing a fleet of Japanese automobiles will pay the Japanese manufacturer in yen, which it purchases with Canadian dollars. In addition, investors speculate on exchange rates, buying those currencies that they believe will increase in value and selling those that they believe will decline in value. Some speculators will borrow money in low-interest rate economies and lend it in high-interest rate economies. Investors and companies exchange currencies to hedge their financial and business risks associated with international trade and financial transactions. Investors also exchange currencies to exploit *arbitrage* opportunities; that is, they buy currencies in markets where they are underpriced and sell them in markets where they are overpriced.

Exchange Rates

Exchange rates denote the number of units of one currency that must be given up for one unit of a second currency. For example, the direct exchange rate of one dollar in terms of the South African rand might be ZAR14, meaning that 14 rand are required to purchase one U.S. dollar. The indirect exchange rate is the inverse of the direct rate. Thus, if one U.S. dollar can be exchanged to purchase 14 South African rand, the direct exchange rate

[12]Bank for International Settlements, "2019 Triennial Central Bank Survey."

of rand for dollars is 14. Correspondingly, the indirect exchange rate of dollars for rand is 0.0729; $0.0729 is the cost of one rand.

Foreign exchange rates are determined by supply and demand conditions. A variety of factors will affect these supply and demand conditions, including:

1. Government policies
2. Supply and demand conditions for commodities in the two countries
3. Income levels in the two countries
4. Interest rates in the two countries
5. The perceived risk of engaging in trade with the two countries

As of 2019, the most commonly traded currencies in spot markets are the U.S. dollar (involved in 89% of all FX transactions), the EU euro (31%), the Japanese yen (17%), the UK pound (13%), and the Australian dollar (7%). The most common trading pair is the EUR/USD, representing approximately 28% of all FX transactions.

Cross Rates

During the Bretton Woods regime following World War II, cross rates referred to the price of a currency other than the U.S. dollar in terms of the price of any other currency besides, again, the U.S. dollar. The Bretton Woods agreement provided that currencies would be pegged to the U.S. dollar. Thus, rates were normally expressed in dollars, although the term cross rate retains its meaning after the Bretton Woods system is largely obsolete. Table 2.7 provides a listing of *cross rates* (the exchange rate or cost of one currency in terms of another) for several different major world currencies along with rates in dollars.

The price of the Swiss franc (CHF) in terms of dollars is given in Table 2.7 to be $1.0217. The price of a U.S. dollar in terms of the UK pound is £0.7676. These two rates imply that the price of a Swiss franc in terms of the UK pound is $1.0217 × £.7676 = 0.7824, consistent with the cross rate given in Table 2.7.

TABLE 2.7 Cross Currency Rates, May 17, 2017

	USD	EUR	JPY	GBP	CHF	AUD	CNY
USD	1	1.1121	0.009	1.3022	1.0217	0.7437	0.1451
EUR	0.8988	1	0.0081	1.1712	0.9184	0.6687	0.1304
JPY	110.803	123.231	1	144.315	113.235	82.411	16.05
GBP	0.7676	0.8536	0.0069	1	0.7842	0.5709	0.1114
CHF	0.9785	1.0883	0.0088	1.2744	1	0.7278	0.1415
AUD	1.344	1.4947	0.0121	1.7507	1.373	1	0.19
CNY	6.8909	7.663	0.0622	8.974	7.0412	5.1248	1

Forward Exchange Rates and Contracts

Exchange transactions can occur in either spot or forward markets. In the spot market, the exchange of one currency for another occurs when the agreement is made. For example, dollars may be exchanged for euros now in an agreement made now. This would be a spot market transaction. In a forward market transaction, the actual exchange of one currency for another actually occurs at a date later than that of the agreement. Thus, traders could agree now on an exchange rate for two currencies at a later date.

A forward exchange contract is a contract specifying delivery of one currency for another at a later date. Thus, a forward exchange contract is simply a contract providing for the exchange of currency at some future date. Forward exchange contracts involve one position in each of two currencies:

1. Long: An investor has a "long" position in that currency that she will accept at the later date.
2. Short: An investor has a "short" position in that currency that she must deliver in the exchange.

Forward exchange contracts will usually involve a third party to act as an intermediary between the exchange parties. This intermediary, generally a bank, is referred to as a market maker. The forward exchange rate, or forward rate, is the number of currency units that must be given up for one unit of a second currency at a later date. This rate may be quite different from the spot rate, which designates the number of currency units that must be exchanged for one unit of another currency now. Forward rates are likely to differ from spot rates simply because supply and demand conditions are likely to change over time. Thus, anticipated changes in the five factors listed above may cause forward rates to differ from spot rates.

There are a number of reasons for participating in forward exchange markets. Many investors participate for the purpose of speculation; they may feel that they know which directions exchange rates will change, and therefore, take positions in contracts enabling them to profit from anticipated changes. Such speculation is often regarded as quite risky. Corporations often take short positions in forward contracts to lock in exchange rates for the dates that they must deliver cash to companies in other countries. Corporations might take long or short positions in contracts to diminish their exchange rate risk. Corporations might take short positions in forward exchange contracts to lock in the exchange rates for foreign-denominated cash flows they anticipate receiving. A smaller group of investors, known as arbitrageurs, purchase forward contracts, and then invest in one of a variety of offsetting securities (including offsetting forward contracts) for the purpose of profiting from market price discrepancies.

Participants in forward exchange markets face a number of risks. Among these are:

1. Rate risk: Exchange rates may change in directions opposite to those anticipated by participants.
2. Credit risk: The other party to the contract may default by not delivering the currency specified in the contract. In many instances, an intermediary such as a large reputable commercial bank may act to ensure that one or both contracting parties will honor their agreements. Dealer reputation is key.

3. Liquidity risk: The market participant may have difficulty obtaining the currency he must deliver; he may be "stuck" with a currency that will be difficult to sell. Again, a number of intermediaries may improve liquidity by trading and making markets for various currencies.
4. Trading system risk: The trading platform, ECN, exchange, and other communication systems are all subject to malfunction or failure.

Because a forward rate specifies the prices investors will pay for currencies at a later date, one might argue that the prevailing forward rate of exchange equals the expected spot rate at the time of currency delivery:

$$F_0 = E[S_t] \tag{2.1}$$

Equation (2.1) states that the forward rate of exchange prevailing at time 0 equals the rate of exchange expected for time t. Speculators will tend to force Equation (2.1) to hold, at least approximately, through their trading activity. However, it is important to remember that the forward rate is based on actual market prices of forward contracts whereas an expected spot rate is merely a prediction by an investor, using prevailing information about prices and other factors such as risk preferences and uncertainty to establish. Hence, there are good reasons for the two values to diverge, including inefficiencies in markets for currencies, commodities, and securities. Nevertheless, there are generally strong relationships between the two for many investors at most times.

Forward markets can involve dealers executing individually negotiated transactions directly with one another through electronic trading systems or by telephone. Futures contracts are standardized transactions with standardized settlement dates on organized exchanges. While forward contracts frequently settle with delivery of the underlying asset, futures contracts typically provide for settlement by cash equivalence.

Futures Markets

Futures contracts differ somewhat from forward contracts in that they are standardized and trade on exchanges, and typically provide for margin (collateral) and marking to the market (changes in margin requirements). That is, futures contracts are traded on exchanges such as the Chicago Board of Trade and the New York Mercantile Exchange. These exchanges and their associated clearing corporations eliminate much of the credit risk and some of the liquidity risk associated with futures contracts. Of course, the rate risk and other forms of risk remain. Standardization of contracts limits the numbers of contracts that can be created, thereby enhancing trader interest in those that do trade. In addition, governments, exchanges, and brokerage houses require that futures traders maintain deposits (margin) to ensure that they will honor their obligations. Often, an investor must vary his margin, depending on the trading environment and the prices of the underlying currencies. This variance of margin, which increases as the value of the investor's obligation increases and decreases as the obligation decreases, is referred to as "marking to the market."

In the United States, a clearing house serves as the intermediary for all futures transactions. Technically, contract participants are all obligated to the clearing house to settle their contracts. Futures trading in the United States, including the trading of FX contracts, is

regulated by the Commodities Futures Trading Commission (CFTC), which functions in a manner similar to the SEC. The forward markets, which generally include only well-known institutions, are self-regulated, and institutions that trade forward contracts are dependent on maintaining their reputations.

Futures and forward contracts are used for hedging and for speculation. Arbitrageurs help maintain an efficient pricing system by seeking situations where contracts and currencies are mispriced relative to one another. Businesses with foreign customers and suppliers frequently use futures contracts to hedge their exchange rate risk. For example, consider a Japanese firm ordering a communications system from an American manufacturer for $5,000,000 payable upon delivery in six months. The Japanese firm is obliged to pay in dollars, but has no control over the price of dollars in yen; thus, it faces the risk that the value of the dollar will increase from its current price of, say, ¥100 per dollar. Since most U.S. firms expect to be paid for their products with dollars, many foreign firms and individuals with this type of exchange rate risk would be discouraged from purchasing American products. However, there are a number of types of derivative securities to control this risk. For example, the Japanese firm could purchase in American markets put options (options to sell) on ¥500,000,000 with an exercise price of $0.01 per yen. This purchase of puts would guarantee the Japanese firm a minimum selling price for ¥500,000,000 needed to purchase $5,000,000. If the price of dollars were to decline, the Japanese would simply purchase the $5,000,000 in cash markets for fewer yen.

Currency Trading

Professional currency traders participate in interbank and interdealer markets as well as exchange markets for various types of contracts on currencies. The most important currency trading centers are located in London, with about 37% of total volume; New York, 18%; Tokyo, 6%; and Frankfurt, Zurich, Hong Kong, and Singapore with somewhat less. Note that the dispersion of markets implies that forex markets operate around the clock. Transactions can be executed between any two counterparties that agree to trade via the telephone or electronic network or trade through an exchange. To a large extent, currency trading is decentralized. Dealers can post or advertise their exchange rate bids and offers using an information or distribution network, such as Reuters and Nex Group.

Many OTC trades proceed through a three-step process similar to the following:

1. A trader communicates by phone or electronically the currency pair and the amount he would like to trade with a given counterpart in trade.
2. The counterparty responds with both a bid and an ask quotation.
3. The trader responds to the bid and ask quotations by buying, selling, or passing on both.

The transaction is then executed if the trader agrees to buy or sell.

Interdealer Brokers and Electronic Broking

The interbank forex markets comprise transactions directly between banks and through electronic brokering platforms. Interdealer brokers facilitate many of these transactions

through highly developed trading platforms. The largest of these brokers, the UK-based NEX Group, formerly known as ICAP Plc, is very active in both voice and electronic markets, averaging over $1.5 trillion daily in all of its brokering services.

FX dealers and certain other financial institutions are permitted to participate in the major electronic broking services. These services provide quotes, permit entry of quotes, and trade execution services. The largest of these with respect to forex trading is EBS BrokerTec, which was created as Electronic Broking Services, a joint venture of several of the world's major foreign exchange market−making banks, and then ultimately acquired by BrokerTec in 2014 to form EBS BrokerTec. It provides quotations, data and news feeds, analytical services and handles close to $200 billion daily in spot FX transactions as well as contracts for several commodities. A major competitor is Reuters Eikon. These services permit straight-through processing, improving speed of transactions and reduced errors.

Futures and Other FX Markets

Currency and interest rate futures are traded on numerous exchanges throughout the world. Among the U.S. exchanges listing currency futures are the International Monetary Market (IMM), a subsidiary of the Chicago Mercantile Exchange (CME), NASDAQ OMX (formerly the Philadelphia Board of Trade (PBOT)), the New York Board of Trade, and the Intercontinental Exchange (ICE). Non-U.S. exchanges trading currencies and currency futures include the London International Financial Futures Exchange (LIFFE, owned by the Intercontinental Exchange Group), Euronext Paris, Eurex Frankfurt, ICE Futures Singapore, and the Tokyo Financial Exchange (TFX).

Orders, Quotes, and Spreads

As noted above, the bid price is the maximum price a dealer is willing to pay for a currency and the ask or offer price is the minimum that she is willing to accept. Dealers typically quote both bid and offer prices, with offer prices normally somewhat higher. The difference between bid and offer prices is called the spread. Narrower spreads are usually seen in more competitive and efficient markets.

An example of a pair of bid and offer quotations on a pair of currencies, say euro/dollar, might be 0.8221/26. Here, the dealer is specifying a bid of EUR 0.8221 per dollar and an offer rate of EUR 0.8226. The spread in this example is 0.0005, or 5 basis points, sometimes called 5 pips. Interbank transactions on the order of $1 million typically quote spreads around 1 to 2 basis points while consumer exchange transactions spreads can be as high as 1000 basis points.

A variety of types of quotation types are used in the trading of currencies. Currency traders must become familiar also with the way that currencies are quoted in the particular environment in which they work. When a currency pair is quoted, the first currency or the denominator currency in the pair is typically called the base currency and the second or numerator currency is the counter or quote currency. In most quotations, the U.S. dollar is the base currency, and quotes are expressed in units of counter currency per USD (e.g., JPY/USD or CAD/USD). The normal exceptions to this convention are quotes involving the euro, the pound sterling, and the Australian dollar; these three are quoted as dollars

per foreign currency. Other quotation conventions are as follows from the perspective of a U.S. trader:

1. Direct quotations (home price of foreign currency, e.g., \$1.8000/£)
2. Indirect quotations (foreign price of home currency, e.g., £0.5556/\$)
3. European quotations (price of USD, e.g., CHF1.8000/\$)
4. American quotations (price of foreign currency, \$0.5556/CHF). European quotes are more common except when trades involve the following currencies: $£_{UK}$, $\$_{AU}$, and $\$_{NZ}$.
5. Outright quotations are the price of a currency in the forward market quoted in terms of a second currency.
6. Point (or swap) quotations are quotations for a currency in forward markets in terms of the premium or discount relative to the spot price. Generally forward spreads exceed spot spreads due to higher uncertainty in forward markets.
7. Forward quotations are often made in percentage terms as follows:
   a. Forward premium or discount as a percent per annum (European terms):

$$[(\text{Spot rate} - \text{Forward rate}) \div \text{Forward rate}] \times 1200/n$$

   b. Forward premium or discount as a percent per annum (American terms):

$$[(\text{Forward rate} - \text{Spot rate}) \div \text{Forward rate}] \times 1200/n$$

8. Cross rates refer to quotations not involving the dollar. These quotations are primarily an artifact of the Bretton Woods system.
9. Eurodollar futures quotation: Stated as an index of the three-month LIBOR:

$$F = 100 - \text{LIBOR}$$

With the exception of the Japanese yen, currency quotations are expressed to four digits. As we mentioned above, the last of these four digits is referred to as a basis point, pip, or percentage in point. The yen is quoted in hundredths (e.g., JPY/USD = ¥120.07/\$).

2.9 CRYPTOCURRENCY AND BITCOIN TRADING

Cryptocurrency and altcoin refer to digital currency in which encryption (cryptography) techniques are used to regulate its creation and supply and to verify its transfer among users. *Bitcoin*, the best known of cryptocurrencies, is a peer-to-peer (P2P) payment network and digital currency created in open-source C++ programming code. Bitcoin does not rely on a central monetary authority to issue, monitor, facilitate, verify, and approve transactions. Instead, Bitcoin is underwritten by a P2P network akin to file-sharing services like BitTorrent (a protocol supporting the practice of P2P file sharing). The network and payment systems are based on an open-source protocol that uses a public but anonymous transaction log. As of February 2021, over 18.6 million bitcoins have been issued out of 21 million planned, with an approximate dollar value of nearly USD1 trillion at more than USD50,000 per coin. Bitcoin surged to more than USD60,000 in April 2021. The rate of increase in the number of bitcoin available to trade has been fairly uniform since its 2009 inception.

Bitcoin came into being in 2009, due to its pseudonymous developer Satoshi Nakamoto. Units of bitcoin (BTC) are created by evidence of *forced work*, also known as mining. Bitcoins are awarded in blocks of up to BTC50 to bitcoin "miners" for solving increasingly difficult proof-of-work computational problems related to public and private keys (discussed below), which are used to confirm transactions and prevent double-spending. Essentially, miners receive bitcoin in exchange for verifying the legitimacy of a block of other bitcoin users' transactions; miners provide the essential service necessary for the accounting of bitcoin.

Generally, one owns money (for example, cash or coins) by being in physical possession of it. One "owns" bitcoin by being able to execute transactions with it; bitcoins exist only as records. Transactions by bitcoin are executed by *hashing*, an update to a public transaction log called a blockchain. Payments are made to and from bitcoin addresses (public keys), which are associated with "wallets," which might be as simple as a scrap of paper with a public key written on it or a webpage requiring a user ID and password. Bitcoin addresses are 33-character alphanumeric strings, such as, for example, *15rDbFcbedaDsYUIerBppokCh8XbbstDfi*. The log to which this address is added is the blockchain, the complete listing of previous bitcoin transactions. A transaction derives from two keys:

1. A *private key*, a 64-character random or difficult-to-guess alphanumeric string analogous to an email password along with a digital signature that is, as was just discussed, derived (hashed) from the private key but that cannot be used to derive or decode the private key.
2. A *public key*, as described above, a 33-character alphanumeric string derived from the private key based on a cryptographic Elliptic Curve Digital Signature Algorithm used to ensure that funds can be spent only by their rightful owners.[13] The public key might be analogous to an email address; everyone knows the address, but only the owner can access its contents or send a message from it. The signature associated with the public key can be used to verify the identity of the sender but cannot be duplicated by anyone without access to the private key.

If the signature can be matched to the public key by computers in the network, the transaction is validated. Essentially, the matching ensures that the bitcoin exists in the wallet and has not been used before. Thus, all transactions are cleared by a database housed on user computers known as *nodes* and verified by miners. Any fraudulent change to any part of the record, represented by a 64-character "hash," will necessitate changes to every other part of the blockchain, also represented by hashes, and on every record maintained by every individual participant in the system. Thus, the blockchain itself protects the integrity of each hash or block in the chain. Node records are publicly available and must be in agreement to validate transactions.

Bitcoin and other cryptocurrencies such as Ethereum and Litecoin are traded in numerous virtual venues including Coinbase and Binance. Such marketplaces provide a variety

[13]An encryption algorithm simply makes a message difficult to decipher without a code. A simple encryption code might change the letter A to the number 1, B to 2, and so on. In this case, the code is quite simple. The Elliptic Curve Digital Signature Algorithm is a more sophisticated procedure based on equations for elliptic curves, which are often used in random number generators.

of services to simplify the ownership and trading processes, including trading platforms such as Coinbase Pro and custodial wallets, which are often insured.[14] Bitcoin activity in near real-time can be viewed at numerous online sites, including https://blockchain.explorer and https://www.blockchain.com/btc/unconfirmed-transactions. Futures contracts on bitcoin are traded on the Cboe Futures Exchange, and bitcoin futures and options are traded on the Chicago Mercantile Exchange.

2.10 COMMODITIES AND FUTURES

Commodities

A commodity market trades in raw or primary products rather than manufactured products. *Soft commodities* are agricultural products such as wheat, livestock, coffee, cocoa, and sugar. *Hard commodities* are mined or extracted, such as gold, rubber, natural gas, and oil. More recently developed commodities markets include those for emissions, electricity, and cell phone minutes. Well-developed commodity markets require that the commodities be well-defined and standardized. For example, one ounce of gold of a given purity should be the same regardless of who owns or sells it. Commodity markets can include physical trading of the actual commodity or a variety of derivative contracts including spot contracts, forwards, futures, options on futures and swap contracts. Trading in futures contracts predates by centuries stock markets; there is evidence of trading in rice futures in China dating to 6000 B.C.E.

Forward and Futures Contracts

A *forward contract* represents an agreement providing for delivery of a given quantity of an asset at a later date at a given price. Essentially, a forward contract provides for the seller (an entity taking the short position on the asset) to deliver the specified asset on a settlement date (or month) to a purchaser (an entity taking the long position on the asset). The actual exchange of the asset for cash occurs at settlement, a date subsequent to the date of origination of the forward contract. A large number of forward transactions involve direct negotiations between banks, brokerage firms, and other financial institutions.

A *futures contract* also represents an agreement specifying delivery of given quantity of an asset on a settlement date at a given price. However, the futures contract differs from a forward contract in several important ways. First, a forward contract is created by its long and short participants according to whatever terms they agree to. A futures contract is created by a *clearing house* which acts as a middleman between all contract participants. The clearing house serves as counterparty on all transactions which effectively eliminates default risk. A futures contract is traded on an exchange. To facilitate trading and to

[14]Although wallets provided by Coinbase and other exchanges are convenient to use, users should be aware that private keys associated with them are usually owned and controlled by the exchanges, but many exchanges permit users to move their cryptocurrencies to their own private (hard) wallets. U.S. users will need to provide personal details needed for tax reporting.

ensure that a reasonably large number of investors will be interested in the futures contract, the futures contract is standardized with respect to the exact quantity and nature of the asset to be delivered along with the settlement date. To ensure that both parties to the futures contract will honor their commitments, futures traders are required to post *margin*, which is, in effect, collateral required by the brokerage firm. Furthermore, futures contracts often provide for *marking to the market*, which involves daily re-computations of the margin requirement based on updated asset value.

Exchanges and clearing houses will clearly specify terms of futures contracts. First, the asset, its quality and grade are specified in the contract, along with the size of the contract (e.g., weight, volume, etc.) and the asset delivery terms. Although physical delivery of the asset is rare because most contracts are *closed out* (canceled by offsetting trades) prior to maturity or provide for the delivery of cash equivalence (*cash settlement*), these terms are still important. Contracts will also provide for a settlement date or delivery month.

There are a number of important exchanges that trade futures contracts, including the Chicago Board of Trade, the Chicago Mercantile Exchange and the New York Mercantile Exchange. See Table 2.1 in Section 2.2 for a listing of major U.S. futures exchanges. In addition to these markets, the London Metal Exchange, the Tokyo Commodity Exchange, and OTC markets are very active. Futures markets in the U.S. are tightly regulated and overseen by the Commodities Futures Trading Commission (CFTC), the National Futures Association (NFA), and to a lesser extent, the Securities Exchange Commission (SEC).

Futures contracts have been traded for many centuries on many types of commodities, including agricultural products such as corn, wheat, and pork bellies, metals such as gold and copper, minerals such as diamonds, petroleum, and phosphates, as well as paper and timber. Futures contracts are traded on an assortment of financial securities such as treasury bonds, equity market indices, and interest rates. As discussed in previous sections, futures contracts also trade on a large number of different currencies.

Futures Contracts and Business Risk

Although futures markets allow investors to speculate and engage in arbitrage, one of their most important functions is to enable businesses to control risk. Consider the following simple example involving a wheat farmer and a cereal manufacturer. Suppose that the manufacturer needs to purchase wheat in three months for cereal production and the farmer, who is just now planting seed wishes to sell wheat at harvest in three months. Since the market price of wheat is likely to be affected by general economic and trade conditions, weather, and many other factors beyond the control of the two trading partners, neither knows what the price of wheat will be in three months. Hence, the manufacturer faces uncertainty with respect to cereal production costs and the farmer cannot know what revenues he will derive from the sale of his produce. Such uncertainty clearly affects the abilities of the trading partners to make appropriate business decisions. In fact, in the face of such uncertainties, businesses often scale back their levels of operations to avoid devastating losses. It would seem simple for the farmer and the cereal manufacturer to agree on a transaction price of wheat in advance. In fact, this is exactly what the futures market allows trading partners to do; the party wishing to lock in a purchase price for

wheat takes a long position in a futures contract and the party wishing to lock in a selling price takes a short position in the futures contract. The futures contract, in effect, obliges each of its participants to transact for wheat at the agreed upon price (settlement price) on the settlement date. However, since the manufacturer never knows exactly whose wheat it will purchase, and the farmer never knows exactly who will be the end user of his wheat, they simply take positions in futures contracts with seemingly anonymous counterparties. Actually, the clearing house is the counterparty for each futures transaction. Although the farmer actually sells wheat at the prevailing market price at harvest, his short position in the futures contract enables to him offset decreases (increases) in the market price of wheat with gains (losses) in his short position in the futures contract. Similarly, the cereal manufacturer purchases wheat at the prevailing market price at harvest, though its long position in the futures contract enables to him offset increases (decreases) in the market price of wheat with gains (losses) in its long position in the futures contract.

2.11 SWAPS AND SWAP EXECUTION FACILITIES

A *swap contract* provides for the exchange of one set of cash flows for another set of cash flows. The amounts of these cash flows are usually tied to values associated with other assets or portfolios but could be tied to practically anything that can be quantified and observed. Many types of swap contracts are custom made (bespoke) and flexible, are often outstanding over longer time frames, and are frequently illiquid. The latter two characteristics present important problems for counterparties seeking to limit risks. Most swap contracts are created and standardized in some respects using a master template provided by the International Swaps and Derivatives Association (ISDA; see Chapter 5).

Swap contracts are regularly specified for interest rates, commodities, currencies, debt and equity securities, and many other types of assets. Swap contracts have a number of uses. For instance, swap contracts enable financial market participants to synthesize other securities that are either unavailable in securities markets or are unfavorably priced. When desired securities or security combinations are unavailable or unfavorably priced, a swap contract can be customized to replicate that security's or combination's cash flow structure. Swap contracts can also be used to obtain exposure to a particular asset or market without having to commit or lock up funds in an investment.

Examples of Swap Contract Types

A *total return swap* is a tradable contract that provides for one party to make a payment based on the total economic performance of a specified asset in exchange for some other fixed or variable cash flow. That is, in the context here, the payments between contracting parties to a total return swap are based on changes in the market valuation or rate related to a specific credit instrument, irrespective of whether a credit event has occurred. A total return swap in effect provides for exchanging an obligation to pay interest at a specified fixed or floating rate for payments representing the total return on a loan (interest and

principal value changes) of a specified amount. Interpreted broadly, other types of swap contracts discussed here might also be considered total return swaps.

Interest rate swaps enable counterparties to swap streams of cash flows associated with two or more different debt instruments or interest rates. For example, a floating-to-fixed swap enables one counterparty to make a payment based on a floating interest rate such as the SOFR (Secured Overnight Financing Rate, a reference rate based on overnight repurchase agreement [repo] transactions) in exchange for a payment based on a fixed rate. The payments exchanged between contracting parties to this interest rate swap are based on changes in the market valuation or rate related to the floating rate credit instrument.

For a more specific example of a "plain vanilla interest rate swap," suppose that a bank enters into a one-year agreement in which the client pays the SOFR on the notional amount of $1,000,000 in exchange for a fixed rate of 5%. Suppose that after one year, the SOFR is 4%. The payment is netted at the end of the swap term with the bank making a payment of $10,000 = ($1,000,000 × [5%−4%]) to the client.[15] In effect, the client has swapped a variable interest payment for a fixed rate payment at 5%.

A *credit default swap* (CDS) provides for one party to pay a fixed premium or series of premiums in return for protection against a specified credit event (e.g., default) or events. A CDS might be used by a lender to arrange for a counterparty to reimburse him in the event of a borrower default. To provide this credit enhancement or insurance, the counterparty accepts a premium or series of premiums for acting as the counterparty.

A CDS can resolve for a lender much credit uncertainty and usually allows for default recovery much faster than court bankruptcy proceedings. The resolution of a borrower default in bankruptcy is never a simple matter. Defaults, debt restructurings, and bankruptcies usually take years to resolve and are usually very costly. The ISDA (discussed in Chapter 5) process speeds up and governs the way borrower defaults are handled with respect to the many types of CDS contracts. Essentially, the ISDA provides for clear definitions of credit events, the declaration of a default, and the setting and design of a bond or debt auction process. This auction can provide for the physical settlement of bonds involved in the default, in which impaired bonds can be liquidated and prices can be discovered. This auction process provides the basis for an appropriate cash flow from short to long CDS position holders and allows for much faster resolution for lenders than court bankruptcy proceedings.

An *equity swap* is a contract providing for the delivery of cash flows associated with shares of equity or an equity index in exchange for the cash flows associated with another asset such as a debt or index instrument. For example, an investor wishing to relieve herself of risk associated with shares that she is currently holding, without selling and exposing herself to capital gains tax liability under some tax regimes, might agree to deliver to another investor the cash flows (dividends and capital gains for a specified period) associated with her shares. The second investor, in turn, agrees to deliver cash flows associated with a treasury bond to the stock investor. Such equity swaps are used to exploit apparent mispricing in equity markets, to manage risks associated with domestic or foreign equity investment, to circumvent dividend withholding tax requirements in foreign countries, to

[15]For a period containing a fraction of a year, the swap payoff would be (Notional Amount × [#Days in Contract/360]) × (Fixed Rate – Market Rate) = $1,000,000 × 360/360 × (.05−.04).

speculate in foreign equity markets when direct ownership is not permitted, and to avoid tax events or certain corporate transaction reporting requirements.

Swap Execution Facilities

A *swap execution facility* (SEF) is a registered electronic trading system that enables OTC swaps traders to provide and obtain quotes and execute swap transactions by accepting bids and offers. An SEF is in many respects similar to a formal exchange, although the term refers to any distributed group of trading systems approved for swaps but not regulated as an exchange. SEFs are intended to provide institutional traders enhanced efficiency and transparency to swaps markets along with reduced systemic risk to the overall financial system. Traditionally (before the Financial Crisis of 2008), most OTC swaps were privately negotiated and traded by financial institutions over the telephone. Such trades were usually cleared bilaterally between trade counterparties.

An SEF maintains Central Limit Order Books (CLOBs; see Section 1.6 and Chapter 4), which collect quotes from market makers to be disseminated to members. An institution (sometimes called a market taker) seeking to accept one of the orders in the CLOB can execute a transaction at the best available price. Many SEFs promote pretrade transparency by maintaining a *request for quote* (RFQ) system. Here, the market taker submits an RFQ (or request for quote stream) to at least three swap market makers, often with assistance of the SEF. Market makers provide quotes, which may or may not be supplemented by the CLOB. Then, if the market taker accepts a quote, the trade is executed through the SEF and reported in compliance with established facility rules to enhance competition, fairness, and transparency. Traders are monitored for adherence to rules, and most trades are centrally cleared to standardize and mitigate fragmented counterparty risk. Once a swap contract becomes traded on a facility, SEFs continue to provide quotes and execution services for that contract, enhancing liquidity and transparency for swaps market participants.

The intermediation by an SEF between two counterparties in a swap agreement is intended to allow counterparty risk shifting by the two counterparties to a trustworthy institution that can take both sides of the transaction. The fairly large number of SEFs, such as those managed by TP ICAP, ICE Swap Trade, Tradeweb, and Bloomberg, are regulated by the SEC (for security-based swaps) and the CFTC. The CFTC maintains a list of registered facilities at https://sirt.cftc.gov/SIRT/SIRT.aspx?Topic=SwapExecutionFacilities. These facilities are a product of Title VII of the 2011 U.S. Dodd-Frank Financial Reform Act (see Chapter 5), which mandates most OTC swaps to be cleared and traded through such facilities if they are not listed on a designated contract market such as an exchange. As will be discussed in the next chapter, Title VII of the Dodd-Frank Act was largely the result of the relative success after the 2008 meltdown of Lehman Brothers in recovering Lehman positions that had cleared with the LCH.Clearnet Group's clearinghouse relative to positions that had not cleared through a clearinghouse. There are exceptions to regulatory facility clearing and trading requirements, such as for highly illiquid and bespoke swap contracts that tend not to be made available to trade.

Most U.S. swap transactions are now executed and cleared by swap execution facilities. Financial institutions obtain authorization to serve as swap execution facilities by meeting

requirements set forth by Dodd-Frank, the SEC, and the CFTC and by submitting a formal application to the SEC.

Additional Reading

Harris (2003), Chapters 1 to 8, provides a nicely detailed introduction to securities and markets and Chapters 13 to 18 will be very useful for understanding market mechanisms and trading institutions. Schwartz (1988) provides a somewhat dated but very readable discussion focused on equity markets in Chapters 1 to 15. The NYSE website (http://www.nyse.com/), through its "About Us" link, offered the latest developments on trading NYSE securities. Other exchanges have similar sites. The World Federation of Exchanges provided much useful information and data concerning exchanges at http://www.world-exchanges.org and should prove useful in any effort to update exchange-related data. White (2013) introduces and reviews the credit rating industry and its characteristics in a very readable format. Hull (2018) provides a useful introduction to swap contracts in Chapter 7, as well as a very comprehensive discussion of derivatives throughout.

References

Bakos, Y., Lucas, H. C., Jr., Simon, G., Viswanathan, S., Oh, W., & Weber, B. (2005). The impact of electronic commerce on the retail brokerage industry. *Information Systems Research*, *16*(4), 352–371.

Baldauf, M., & Mollner, J. (2020). Trading in fragmented markets. *Journal of Financial and Quantitative Analysis*, *56*(1), 93–121.

Bank for International Settlements. (2019, December). *Triennial central bank survey: Foreign exchange turnover in April 2019* <http://www.bis.org/publ/rpfx19fx.pdf>.

Battalio, R., Greene, J., & Jennings, J. (1997). Do competing specialists and preferencing dealers affect market quality? *Review of Financial Studies*, *10*, 969–993.

Battalio, R. J., Hatch, B., & Jennings, R. (2004). Does a national market system exist for U.S. exchange-listed equity options?. *Journal of Finance*, *59*, 933–962.

Breedon, F., & Holland, A. (1997). *Electronic versus open outcry markets: The case of the bund futures contract.* Working paper. London: Bank of England.

Cavaletti, C. (1997). Commission rate ride. *Futures*, *26*(12), 68–70.

Cboe. (2021). *U.S. equities market volume summary 2021*. Available at <https://www.cboe.com/us/equities/market_share/> Accessed 11.09.21.

Chung, K. H., & Chuwonganant, C. (2021). *COVID-19 and the stock market: Liquidity, pricing efficiency, and trading*. Unpublished working paper. Albany, NY: State University of New York. Available at <http://dx.doi.org/10.2139/ssrn.3798625> Accessed 30.06.21.

Frino, A., Hill, A. M., Jarnecic, E., & Aitken, M. (2004). The impact of electronic trading on bid-ask spreads: Evidence from futures markets in Hong Kong, London and Sydney. *Journal of Futures Markets*, *24*, 675–696.

Harris, L. (2003). *Trading and exchanges:* Market microstructure for practitioners. Oxford: Oxford University Press.

Hull, J. C. (2018). *Options, futures and other derivatives* (10th ed.). New York: Pearson.

Klein, A. (2004). Credit raters' power leads to abuses, some borrowers say. *Washington Post A01* <http://www.washingtonpost.com/wp-dyn/articles/A8032-2004Nov23.html>.

Options Clearing Corporation. (2021). *Options value by exchange–2020*. Available at <https://www.theocc.com/Market-Data/Market-Data-Reports/Volume-and-Open-Interest/Volume-by-Exchange> Accessed 10.10.21.

Pirrong, S. C. (1996). Liquidity and depth on open outcry and automated exchanges: A comparison of the LIFFE and DTB bund contracts. *Journal of Futures Markets*, *15*(5), 519–543.

Raj, P. (2017). *Origins of impersonal markets in commercial and communication revolutions of Europe*. Unpublished working paper. Chicago, IL: Stigler Center for the Study of the Economy and the State, University of Chicago. Available at <https://research.chicagobooth.edu/-/media/research/stigler/pdfs/workingpapers/11emergenceofimpersonalmarketsprateekraj.pdf> Accessed 30.08.21.

Rosov, S. (2016). *Payment for order flow in the United Kingdom: Internalisation, retail trading, trade-through protection, and implications for market structure*. CFA Institute. <http://www.cfapubs.org/doi/sum/10.2469/ccb.v2016.n8.1>.

Sarkar, A., & Tozzi, M. (1998, January). Electronic trading on futures exchanges. Federal Reserve Bank of New York. *Current Issues in Economics and Finance*, 4(1), 1-6.

Schwartz, R. A. (1988). *Equity markets: Structure, trading and performance*. New York: Harper and Row Publishers.

Securities and Exchange Commission. (1998, December). *Regulation of exchanges and alternative trading systems*. Release Number 34-40760.

Securities and Exchange Commission. (2021a). *Alternative trading systems with Form ATS on file with the SEC as of October 31, 2021*. <https://www.sec.gov/files/data/alternative-trading-system-ats-list/atslist103121.pdf>.

Securities and Exchange Commission. (2021b). *Self-regulatory organization rulemaking*. Available at <https://www.sec.gov/rules/sro.shtml> Accessed 10.10.21.

Shyy, G., & Lee, J. (1995). Price transmission and information asymmetry in bund futures markets: LIFFE vs. DTB. *Journal of Futures Markets, 15*, 87–99.

Solnick, B. (1991). *International investments* (2nd ed.). Reading, MA: Addison-Wesley.

Stigler, G. (1961). The economics of information. *Journal of Political Economy, 49*, 213–225.

Weber, J. (2007). Behind the burst of the bourses. *Business Week, 59*.

White, L. J. (2013). Credit rating agencies: An overview. *Annual Review of Financial Economics, 5*, 93–122.

Xia, H. (February 2014). Can investor-paid credit rating agencies improve the quality of issuer-paid rating agencies? *Journal of Financial Economics, 111*(2), 450–468.

2.12 EXERCISES

1. For a 20-year period beginning in 1986, the NYSE was regularly criticized for making less use of electronic and computer technology than most of its competitors around the world, even less than exchanges in emerging and developing countries and than exchanges just starting business. The NYSE would argue that its specialist system required more "face-to-face" contact than other nonspecialist exchanges.
   a. Why would the NYSE specialist system require more face-to-face contact than its competitors?
   b. Assume that face-to-face contact hurt investors more through increased trading expenses and reduced trading efficiency than it helped them. That is, assume that the argument that you prepared for part (a) is simply wrong. Why might the NYSE still resist adopting improved electronic and computer technology?
2. Why is the NYSE trading model often referred to as a hybrid model of trading?
3. An ATS (Alternative Trading System) might be defined as an exchange that does not operate as a self-regulatory body. True or false?
4. Might a bond's credit rating imply any information about the liquidity risk or interest rate risk of a bond?
5. Suppose that the spot price of gold is $1100 per ounce. The current one-year riskless rate of interest is 3%.
   a. Provide a forecast for the price of gold in one year based on a relevant expectations hypothesis.
   b. Provide an estimate for the one-year forward price for one ounce of gold.
   c. Justify your use of the expectations hypothesis that you used in your answer to parts a and b.
   d. What would cause the forward price of gold to deviate from your forecast?

6. Suppose that the Swiss franc is currently trading at 2 per U.S. dollar. The franc is expected to be revalued to CHF2.5 per dollar.
   a. Has the franc been devalued relative to the dollar?
   b. Has the dollar been devalued relative to the franc?
   c. What is the percentage anticipated revaluation of the franc relative to the dollar?
   d. What is the percentage anticipated revaluation of the dollar relative to the franc?
7. Assume that $1 will purchase £0.60 and ¥108; that is, one U.S. dollar will purchase 0.6 UK pounds and 108 Japanese yen. Assume that goods in the three countries are identically priced after adjusting for currency exchange rates and that there are no arbitrage opportunities.
   a. What is the value of £1 in ¥?
   b. What is the value of ¥1 in £?
   c. If one ounce of gold costs $1500 in the United States, what is its cost in the UK and in Japan?
8. Why must a forward price equal the spot price on the settlement date of the forward contract? That is, why do many forward prices converge to spot prices at settlement?
9. What are the differences between forward and futures contracts?
10. Equity swaps have been used by investors to reduce their risk in an equity investment without actually selling shares. For example, corporate managers have used the *executive equity swap* to reduce their personal exposure in their shares of their employers' stock. In a well-publicized case involving Autotote Company, at the time a Nasdaq-listed manufacturer of wagering equipment, the CEO, Lorne Weil, arranged a swap contract through Bankers Trust (since taken over by Deutsche Bank) to deliver dividends and any capital gains (which would be negative in the event of a capital loss) associated with Autotote stock in exchange for associated cash flows pegged to the variable interest rate LIOBR. How might Mr. Weil have benefitted from this transaction?
11. Assume that exchange rates of Swiss francs per dollar are 1.6 and 1.8 in the spot and one-year forward markets, respectively. Thus, an investor can transact for 1.6 francs per dollar today and contract to transact for 1.8 francs in one year. Assume that nominal interest rates are 10% in the U.S. and 12% in Switzerland. Demonstrate an arbitrage opportunity. That is, demonstrate how an investor can obtain an arbitrage profit by simultaneously using spot and forward markets to transact for securities and by borrowing and/or lending money. Further assume no transactions costs or barriers on transactions and that interest rates apply to both borrowing (short) and lending (long).

CHAPTER

3

Financial Market Utilities and Securities Trading Support

3.1 SECURITIES TRADING INFRASTRUCTURE

Casual traders are usually well-served by their retail brokerage firms in dealing with the mechanics of trading. Such traders, often making a small number of trades per day, receive most of the services and details that they require from their retail brokers and avoid many of the complexities that professional traders encounter when dealing with the mechanics of trading. Professional traders often encounter more complex trading regulations and technicalities and need to be better informed with respect to the "nuts and bolts" of their trading activities. Even simple issues, which can usually be ignored by casual traders, become more important to professional traders. For example, how do traders take possession of the securities they buy? How do they receive cash for the securities they sell? What is the total cost of derivative exercise or settlement? How are exercise and settlement requirements fulfilled for their derivative transactions? What are the costs and risks associated with all these issues? These trader activities all occur within the *financial market infrastructures*, sometimes referred to as the "plumbing for financial markets," which provide key services that are critical to the smooth and reliable functioning of markets. Individual institutions providing such services often are referred to as *financial market utilities*, which include central counterparties and provide reliable transfer, payments, clearing, and settlement systems. Serious traders need to be familiar with such systems and institutions, especially those that involve clearing firms and clearinghouses, along with their practices and the requirements they face when dealing with them.

We will narrow our discussion in this chapter to focus on securities trading infrastructure and utilities while broadening our discussion to include institutions that provide pretrade, brokerage (execution), and posttrade services, including quotations, data, and prices. Some broader definitions of financial market infrastructure include exchanges (we discussed these in the previous chapter), cloud computing systems, cybersecurity, and various support systems for banks and other depository institutions. Most of these services are beyond the scope of this book. Here, we will introduce quotations systems, followed

DOI: https://doi.org/10.1016/B978-0-323-90955-6.00003-3

© 2023 Elsevier Inc. All rights reserved.

by brokerage operations. We will then emphasize traditional securities market infrastructures such as clearing and settlement systems and central counterparties, along with the other services they provide.

3.2 PRICE AND QUOTATION SYSTEMS

Traders can subscribe to direct price and quote data feeds from any number of the various securities exchanges in the U.S. At the same time, for example, stock traders and traders in certain other markets have to contend with the same securities being traded in multiple venues. That is, U.S. stock and other securities markets are fragmented, which hinders the informative quality of pricing and market efficiency and can hinder the ability of traders to obtain the best possible prices for their transactions. The costs of subscribing to and analyzing multiple data feeds is prohibitive for most traders. The U.S. Congress has made several efforts to legislate reductions in market fragmentation and its harmful effects on investors without hindering market competition. The U.S. Congress has also made efforts to improve competition between markets for securities trading. For example, the Securities Act of 1975 amended the Securities Exchange Act of 1934 to facilitate development of a national market system for securities transactions. With the national system created by the 1975 Act, Congress intended to increase competition, efficiency, and price transparency in securities markets, and to promote *best execution* and order interaction. In this system, national and regional exchanges would list security prices simultaneously on relevant quotation and trade reporting systems. The purpose of this national market system was to ensure that investors would receive the best possible executions for their securities transactions, where the meaning of best execution, in a rather vague manner, seeks to account for the best price given the existing NBBO, execution speed, differences in quantities of securities actually offered and bid, transactions costs and market depths. We will discuss further the concept of best execution in Chapters 4 and 5.

The Securities Industry Automated Corporation (SIAC), fully owned by the NYSE, maintains the primary systems for disseminating NYSE price and quote information to the public. The National Market System's prices and quotations are compiled and disseminated by securities information processors (SIPs). The system operated and maintained by the NYSE is simply called the *Securities Information Processor* (SIP), although a second SIP is operated by Nasdaq.

SIPs link U.S. markets by collecting, consolidating, and compiling all trade prices and all bid/ask quotes along with certain other critical information including highs, lows, and the NBBO from all relevant trading venues into a single continuous data feed. SIPs seek to mitigate the adverse effects of market fragmentation while retaining competition among trading venues, including registered exchanges and certain other markets, which are required to provide quotation and transaction data to a designated SIP. The two U.S. equity SIPs are the *Consolidated Tape Association* (CTA), overseen by the NYSE, and *Unlisted Trading Privileges* (UTP), overseen by Nasdaq. These SIPs combine and disseminate prices and quotes through three SIP networks: Tape A, governed by the CTA, which reports for NYSE securities regardless of where they trade; Tape B, also governed by the CTA, which reports for exchanges, ATSs, and venues other than the NYSE and Nasdaq;

and Tape C, governed by UTP for securities listed on Nasdaq (see Securities and Exchange Commission [2020]).

The CTA's two reporting systems are the Consolidated Tape System (CTS) and the Consolidated Quotations System (CQS). The CTS is a high-speed electronic system that reports transaction prices and volumes for securities on most U.S. equity exchanges and markets, displayed by some data vendors in "Time & Sales Windows." The CQS provides traders with price quotes from the various equity market venues and FINRA and calculates and disseminates the NBBO and certain trading restrictions, mostly related to short sales and market halts, displayed by some vendors in a limit book format. Both CTS and CQS data are combined into network Tapes A, B, and C and sent to subscribers. In addition, data compiled by the SIAC for the CTA is also compiled to the U.S. options SIP, the *Options Price Reporting Authority* (OPRA).

These information processing systems might be regarded to be the centerpieces of the *Intermarket Trading System* (ITS), a general term for the market system arising from 1970s U.S. Congressional efforts to enhance market competition and mitigate the negative effects of fragmentation. The ITS displays quotes from different markets, compiles the NBBO and has linked markets for trade executions and to facilitate investors' access to the best quotes.

Information and data are valuable assets, and sales of these data are important sources of revenue to exchanges and their members. An SIP will charge its many users of data for access to its data. The varying types of users pay different rates for different use patterns and different quantities of data, with typical retail access through brokers costing more on a per query basis (usually, the costs are absorbed by their brokers) than professional user fixed prices, which usually cost more overall for continuous and more detailed access. Many traders access this price and quotations data through various types of vendors as described in Chapter 1 of this book. Media organizations purchase data at yet different rates. To encourage information production and improved markets, the SIP rebates revenues to markets and traders, who regularly contribute and improve price quotes and contribute more data to the systems.

3.3 BROKERAGE OPERATIONS

Securities brokerage firms have linked traders and investors who are widely dispersed throughout the world, enabling them to execute transactions in central physical and electronic venues. Securities brokers traditionally received commissions for their execution services and often provided a variety of other services to their clients. The process of executing securities transactions has become increasingly complex over the years, but brokerage firms continue to appear as the face of trading processes to most investors.

Until 1975, brokerage firms in the United States functioned as a rather exclusive cartel, fixing commissions and offering clients a wide array of services ranging from providing trade executions and margin opportunities to offering advice and management services. Brokerage firms were all said to be full service, and included firms such as Merrill Lynch, E.F. Hutton, Paine Webber, and Smith Barney. The industry was forced as of May 1, 1975, acting on SEC Rule 10b-3 to break its cartel and compete with respect to brokerage commission levels. Numerous discount brokers opened operations, competing against full service brokers by offering trade executions with lower brokerage commissions. Discount

brokers initiating operations in the 1970s included Quick & Reilly and Charles Schwab. The Internet's development in the 1990s and shortly afterwards created opportunities for Internet brokers such as Ameritrade, ScottTrade, Interactive Brokers, and E-Trade to further reduce commissions and execute transactions for clients without actually speaking live with brokerage account representatives.

Investors wishing to open brokerage accounts need to consider their trading needs and interests to properly select their brokers. Investors needing substantial advice and counsel along with other services may find their needs better fulfilled by full-service brokers such as Merrill Lynch, Raymond James, or UBS. In addition, there is some statistical evidence (discussed earlier) suggesting that larger full-service firms better provide for transaction price improvement (executing the client's transaction at a price that is better than the best bid or offering at the time of the transaction). Investors who are comfortable with their trading experience and who expect to execute smaller trades, particularly more frequent smaller trades, might prefer the lower commissions offered by an Internet broker such as ScottTrade or T.D. Ameritrade. These on-line brokers will often offer low flat-rate commissions, sometimes even dropping to zero, particularly when the brokerage firms receive payment for order flow (rebates).

J.D. Power and Associates regularly provides broker quality information based on its customer satisfaction surveys. Important evaluation criteria include commission levels, percentage of orders showing price improvement, frequency of limit order executions, margin and cash interest rates, and customer satisfaction. *Barrons* conducts similar surveys. Such surveys are useful in making broker-selection decisions.

Most stock orders in the United States are placed through stockbrokers who were traditionally compensated with commissions. These commissions are or were typically charged as a fraction of the dollar volume of the transaction and as a function of the number of shares associated with the order, or in many instances, a simple flat commission (recall sources of brokerage revenues in Section 1.7). Sometimes, the broker's commission will be a function of the bid-ask spread; that is, the difference between the lowest price a trader is offering for the security and the highest price an investor is willing to bid for that security. OTC markets and particularly exchanges will provide for trading priority rules, which designate which orders will be filled or the sequence in which orders will be filled. Frequently, priority rules are designated according to transaction price, size, location of origination of order timing, and so on.

Securities Investors Insurance Corporation

The "Paperwork Crisis" of the late 1960s and early 1970s (to be discussed shortly) caused many brokerage firms to fail, at the cost of public confidence in stock markets. In response, the *Securities Investors Protection Corporation* (SIPC) was created by the U.S. Congress as a nonprofit corporation under the Securities Investor Protection Act of 1970 to protect clients from failures of registered securities brokerage firms. In some respects, the SIPC is analogous to the Federal Deposit Insurance Corporation (FDIC) in that it insures investor accounts in U.S. brokerage firms for up to $500,000 in cash and securities, including up to $250,000 in cash (money market mutual funds are considered by the SIPC to be securities), but only in the

event of the failure and liquidation of the brokerage firm. Thus, this coverage is very limited. The SIPC does not provide protection from declines in security values, nor does it provide protection for losses from securities not registered with the SEC or even from fraud unless the brokerage firm fails. Coverage extends to non–U.S. citizens.

The SIPC assesses its members, normally annually at a rate of 0.0015 of net operating revenues as of 2021. These assessments are used to contribute to a default fund, which as of year-end 2020 had reached an overall value of approximately $3.9 billion. The SIPC seeks to ultimately grow this fund to $5 billion. The SIPC also maintains a $2.5 billion line of credit with the U.S. Treasury. From its creation in 1970 to year-end 2020, the SIPC had injected $3.1 billion from its fund to cover investor losses, enabling it to recover approximately $141.8 billion on behalf of 773,000 insured investors (see SIPC [2021]).

Back Office Operations

Most securities brokerage firms will operate front, middle, and back offices, which refer to groups that provide different categories of functions and services. Traditionally, front-office personnel produce revenues by engaging in direct interaction with the broker's clients (client-facing), middle-office personnel directly support front-office staff, and back-office personnel provide support services behind the scenes. Operations or back-office personnel include IT specialists, accountants, operations staff, human resources staff, office managers, customer service representatives, and regulatory compliance staff. As we mentioned in Section 1.2 and will detail shortly in this chapter, crucial back-office activities include trade confirmation, clearing, and settlement functions, almost always conducted with the assistance of clearing institutions, as we will discuss shortly. Back-office activities also include margin calculations, exercise and settlement for derivative securities, records and account maintenance, software and information technology, and data provision and might include efforts related to strategic planning for the brokerage firm. Back-office brokerage personnel often work with clearing firms (especially when their employers do not self-clear) and clearinghouses.

3.4 CLEARING AND SETTLEMENT

The general *clearing* process involves two primary tasks: trade comparison (matching of trades) and settlement (physical delivery of securities and cash or book entry delivery). Clearing refers to the set of activities resulting in settling claims of financial institutions against other financial institutions. The operations department of a financial institution, often referred to as the institution's back office, is responsible for handling or overseeing the clearing and settlement processes for their own transactions.

Clearing Firms and Clearinghouses

Only approximately 5% of over 5000 U.S. brokerage firms are self-clearing; that is, only 5% are capable of clearing trades for themselves. Other brokers, known as correspondent firms, forward their orders to a *clearing firm*. A clearing firm is authorized by a

clearinghouse (defined below) to manage trade comparisons and other back office operations. Leading clearing firms include Pershing, LLC, RBC Correspondent Services, and Southwest Securities, Inc. Most traders who are unaffiliated with self-clearing institutions will clear their transactions through such clearing firms.

A *clearinghouse* clears transactions for a market such as the NYSE and is typically established by that market or members of that market. A clearinghouse facilitates the trade settlement between two clearing firms (also called as member firms or clearing participants) and seeks to ensure that the clearing firms honor their trade settlement obligations. The clearinghouse will typically guarantee the obligations of its member firms, and often require collateral to ensure that settlement obligations are fulfilled. The collateral is pooled into a clearinghouse guarantee fund. The clearinghouse will step into a transaction to be settled by its members and assume the settlement obligations of both counterparties to the transaction, in effect becoming the counterparty to both sides of every transaction, a process known as *novation*. Thus, the clearinghouse, acting as a *central counterparty*, acts as the counterparty for each party to every transaction, and assumes all credit risk associated with each party. Option clearinghouses will handle the process of option exercise.

In the United States, the *National Securities Clearing Corporation* (NSCC, a division of the Depository Trust and Clearing Corporation, described below), the successor to the combined clearing corporations of NYSE, AMEX, and NASD, is the major clearing agent for equity markets. NSCC serves as the clearing agent for these three major equity markets as well as for many bond markets. Its primary facility for clearing is the Securities Industry Automated Corporation (SIAC), which maintains the computer systems for clearing. Clearinghouses clear and settle trades for most options, futures, and swap markets. Most U.S. options exchange transactions are cleared through and guaranteed by the Options Clearing Corporation (OCC). The OCC clears transactions for options and futures on equity instruments, stock index contracts, currencies, and certain interest rate contracts as well as a number of contracts in other markets.

Trade Confirmation and Comparison

Confirmation is the first step of the clearing process, occurring after the trade reports are sent to traders' clearing agents (if relevant) who forward them to the central clearing agent (if the market centrally clears) by both counterparties in the trade. When trades are executed, buyers and sellers record trade details. Brokers and dealers receive confirmations that the trade has been executed and pass on details of the confirmation to clients. The typical confirmation document received by the client reports the security's name and CUSIP number (the security's nine-character alphanumeric identifier issued by the Committee on Uniform Security Identification Procedures), the number of units traded, the security price, and the broker commission and other fees, along with trade and settlement dates.

Trade comparison is the second step in the clearing process. Comparison matches counterparties in transactions. Trades are compared and are said to be cleared when the counterparties' records are identical. This happens for the vast majority of trades. *Out-trades* in futures markets or *DKs* (*don't knows*) in other markets, which are trade reports with discrepancies resulting from recording errors, misunderstanding and fraud, are sent back to traders to resolve or reconcile.

Netting

The number of securities transactions that occurs each day is huge, requiring some sort of process to simplify the process of changing title to securities and moving corresponding cash proceeds between accounts. *Netting* is the simplification process used by clearing firms, and is one of the most important functions of the clearing process, and of the NSCC. Netting is the process of adding all of an institution's purchases of each security, adding the sales of each security, deducting sells from buys to determine the net change in holdings of that security for the institution and computing the net cash flows associated with all transactions. The NSCC is able to reduce the total number of trades requiring an exchange of money by about 98% through the netting process, where millions of transactions are reduced to each clearing member receiving net debit and credit notifications for cash and for each traded security. The NSCC uses an automated system through the Securities Industry Automation Corporation (SIAC) to "net down" or reduce the number of trading obligations that require financial settlement.[1] Since most brokerage firms execute large numbers of both buy and sell transactions for the securities that they trade, the netting down process results in only about 3% of the total reported matched or offsetting transactions actually having to be settled between financial institutions. At the end of the netting process, the NSCC delivers to each brokerage firm settlement instructions.

Custody and Trade Settlement

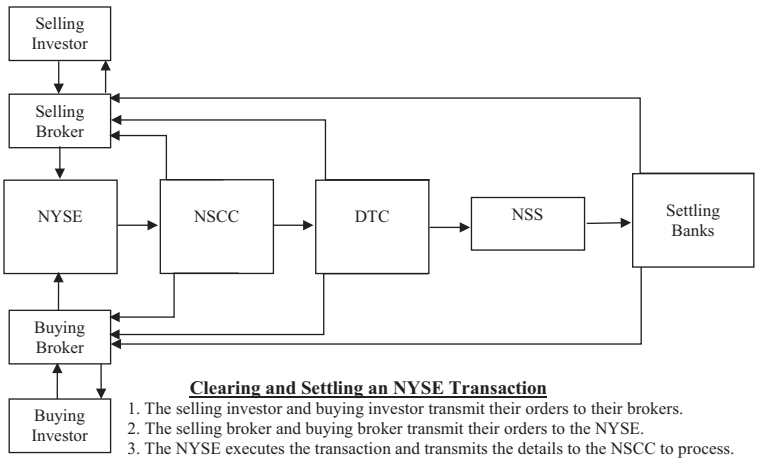
Trade settlement occurs when buyers receive their securities and when sellers receive payment for their securities. Stock ownership in most cases is evidenced by a certificate. The *Depository Trust Company*, a subsidiary of the Depository Trust and Clearing Corporation, holds stock certificates of member firms, registering them in member names and maintaining computerized records of ownership and transfers.[2] The Depository Trust Company and the National Securities Clearing Corporation, the two major units of the DTCC, were both founded as a result of paperwork crises in the 1960s and 1970s when messengers were delivering paper securities certificates by hand on behalf of brokerage firms and settling transactions on paper. As of 2018, the DTCC was holding in custody approximately $52.2 trillion in securities and processed a daily average of over 1.4 million transactions during the year. Overall, the DTCC settled more than $2.329 quadrillion in securities transactions in 2020, including equities, fixed income instruments, mutual funds, insurance products, and so on. Equity securities are held in street name, meaning that securities are held in the names of brokers (nominee names), who, in turn, maintain their own records of ownership in client

[1]More generally, the SIAC, a subsidiary of the NYSE, runs the computer systems and communications networks on behalf of the NYSE and performs clearing functions on behalf of clearing firms.

[2]Technically, securities are held in custody by a DTC partnership affiliate known as Cede & Co., an acronym for Central Depository. Shares are immobilized in the nominee's name, Cede & Co. In the case of direct registration systems such as dividend reinvestment plans, the issuing firm will place the security certificates with a transfer agent, which is normally a bank with an affiliate structured for this purpose.

accounts.[3] Settlement of a trade is completed when the DTCC transfers the ownership of the shares from the selling firm to the buying firm in its automated book-entry recordkeeping system and transfers money between firms with net credits and net debits. Firms that have net credit after end-of-day netting are owed more by other brokers than they owe. As depicted in Figure 3.1, the *National Securities Clearing Corporation* (NSCC) transfers money to banks of these net credit firms from accounts of brokers with net debits through the *Fedwire* system. At present, the SEC requires that settlement occur within two days (T+2) after most transactions. In addition, the DTCC provides cost basis information and other information services. The DTCC provides clearing services at a very low cost, averaging approximately $0.018 per trade.

Modern settlement of trades normally takes place in electronic book entry form. However, during most of the 19th century, each NYSE transaction required delivery of paper certificates and checks and had to be settled within one day (T+1), creating significant pressures as messengers rushed to make deliveries after transactions. Failure to make timely delivery impaired the performance of counterparties, intensifying contagion risks.



Clearing and Settling an NYSE Transaction

1. The selling investor and buying investor transmit their orders to their brokers.
2. The selling broker and buying broker transmit their orders to the NYSE.
3. The NYSE executes the transaction and transmits the details to the NSCC to process.
4. The NSCC compares, records, and nets trades, sending details back to the brokers and the DTC on net security positions and monetary amounts to be settled.
5. The DTC nets and transfers ownership of securities in its electronic books of member brokers. The DTC sends instructions to the National Settlement Service (NSS), which adjusts settling bank funds accounts for net funds transfers.
6. Settling banks withdraw money from the net buying broker accounts; settling banks add money to the net selling broker accounts.
7. Buying and selling brokers settle their various client accounts accordingly.

FIGURE 3.1 Clearing and Settling an NYSE Transaction[4]

[3]Actually, shareholders can maintain stock custody themselves through physical certificates or, if the company allows, hold their stock through direct registration with the company. These alternative choices significantly diminish the convenience of transaction execution relative to registration in street name.

[4]See a more detailed version of the clearing and settlement process at Securities and Exchange Commission (2015), p. 45, particularly with respect for involvement of entities such as transfer agents and introducing brokers.

In response to these pressures and several financial market panics, the NYSE introduced netting to the clearing processes with the founding of the NYSE Clearinghouse, later to be called the Night Clearing Branch. In 1920, the Stock Clearing Corporation (SCC) was established by the NYSE to serve as the central counterparty for every transaction. The establishment of the SCC provided for novation and meant that market participants needed to have checks and stock certificates delivered daily only to a single central counterparty, the SCC. A *central counterparty*, typically a clearinghouse, serves the role of buyer to every seller and seller to every buyer in each market. The central counterparty through novation drastically simplifies paperwork and delivery relative to bilateral settling (parties to each transaction settle with one another), record-keeping, and securities transfer. The central counterparty also guarantees contractual obligations, thereby reducing the risk of market failure. Members of the clearinghouse, called clearing members, interact directly with the central counterparty both as principal and as agent for their clients, which include brokerage firms, institutional traders, smaller financial institutions, insurance companies, etc.

NYSE market activities continued to expand, and despite the novation innovations implemented by the DTCC, strains on clearing and settlement systems resumed. As mentioned earlier, in the 1960s, securities clearing involved the physical transfer by messenger of paper securities and checks. By 1967, this system became untenable due to the tremendous paperwork backlogs associated with increasing securities market activity. Ultimately, the volume of trading activity led to closure of the NYSE on Wednesdays, shortening of other trading days, and increasing settlement to T+5. These backlogs and trading impairments, collectively known as the *Paperwork Crisis*, diminished investor confidence in the safety of U.S. capital markets and were a contributing factor to more than 100 broker-dealers and acquisitions operating under stress or failing. The *Depository Trust Corporation* (DTC) was set up in 1973 to handle the clearing business for the major stock exchanges. Since 1999, the DTCC has owned the Depository Trust Company and the National Securities Clearing Corporation. The DTCC also owns the Fixed Income Clearing Corporation and Deriv/SERV, LLC. The DTCC is owned by its customers, exchanges, and ECNs that clear through their facilities. The NSCC and the DTCC also require brokerage firms to maintain collateral to ensure that they fulfill their obligations. In Europe, the major central securities depositories are Euroclear (the world's largest clearance and settlement system for internationally traded securities), EuroCCP (currently owned by Cboe Global Markets), and Clearstream (Deutsche Börse Group). The London-based LCH Group (formerly the London Clearing House) provides a variety of clearing services in the United Kingdom.

In addition to producing obvious efficiency gains, central counterparties serve another important function: they improve market resilience by increasing transparency and reducing counterparty risk. For example, the OTC derivatives markets faced unprecedented risk levels and contagion (spread of economic crisis) in September 2008 due partly to the huge number of bilateral (no central counterparty) derivative contracts involving Lehman Brothers, the giant dealer on the brink of bankruptcy. The market's inability to net and settle the roughly 930,000 outstanding bilateral derivatives contracts with Lehman threatened to cascade and melt down the entire financial industry. On the other hand, LCH Swapclear, the clearinghouse for 66,000 Lehman trades with notional value of $9 trillion in five currencies, successfully managed contracts that it served in the wake of the collapse

(Loader, 2020). Within a month, the losses to LCH and its default fund (funded by LCH Swapclear members to enhance default protection) were zero, and the surplus Lehman margin was returned to Lehman's bankruptcy administrators. Thus, the Lehman crisis might have been mitigated with novation and use of a central counterparty for its OTC transactions, due in part to the ability to net down individual trader settlement obligations more easily. Use of a central counterparty spreads and mutualizes risk among market participants, whereas standardizing margin and other trading requirements reduces market abuses and improves transparency through improved monitoring and efficiency of information gathering. Numerous studies, including one by Bernstein, Hughson, and Weidenmier (2019), have found empirical results confirming both risk-reduction effects and share price improvement resulting from use of clearinghouses and central counterparties.

End of Day Settlement

Central settling provides for end-of-day trade settlement of net funds with settling banks. Members of the clearinghouse designate their own settling banks, from which their settlement funds are disbursed and collected. In the U.S., the DTC and NSCC use the National Settlement Service (NSS) provided by the Federal Reserve System and the Fedwire system to settle DTC member funds accounts. Each day, the DTC posts final net-net figures for each clearinghouse member for securities and each settling bank for cash, with the designated settling banks using the NSS to transfer funds among the banks to settle their cash obligations and to directly post debits and credits to bank accounts (see Figure 3.1). Each settlement bank will then transfer funds accordingly to and from its own designated clearinghouse members.

Investors can commit settlement violations if their new trades are not fully covered by settled funds, which include cash balances, incoming cash, available margin, and settled proceeds of fully paid-for securities. Traders should be sure to understand their broker's policies concerning settlement violations such as purchasing a stock through a cash account without settled funds or sufficient settled funds, then selling that stock before paying for its purchase. While such transactions may well be permissible in a margin account with sufficient margin, it is likely that the trader's broker will prohibit such transactions in a cash account.

Margin, Collateral, and Hypothecation

As discussed in Chapter 2, client account risk with brokers is reduced through the imposition of margin requirements on clients. Margin requirements are normally imposed on the purchase of equity and other securities with borrowed funds, short sales, positions in futures contracts, writing of options contracts, etc. These margin requirements are analogous to the posting of returnable collateral on open positions, ensuring that clients fulfill their trading obligations; that is, settlement risk in the market is mitigated through the imposition of margin or collateral requirements, sometimes known as *performance bonds*. As discussed earlier, central counterparties (through novation) further reduce settlement risk by requiring collateral from their members, who are mostly brokers and dealers.

Posted collateral in the clearinghouse default fund (also called the participants' fund) for each member is a function of member trading activity, the volatility of member-traded securities, imbalances between member buying and selling, and the member's financial condition. Posted collateral in the custody of the DTC is maintained by its Collateral Monitor, whereas the OCC in 2016 held approximately $100 billion in collateral posted by its members. If securities are used as collateral in default funds, the DTC will reduce their value as collateral relative to cash; that is, the securities are given a *haircut* (the difference between the market value of a security and its value as collateral).

Most clearinghouses require of their members minimum collateral calculated on at least a daily basis. Similarly, brokers require a variation (maintenance or mark to market) margin of their clients, typically adjusted daily depending on the risks associated with the client accounts. In some scenarios, especially when markets are stressed, brokers and central counterparties will require *additional* margin from their clients or members to account for widening of spreads and reduced liquidity. Broker collateral with clearinghouses will be netted across client accounts and positions and will also vary daily. Both initial margin and variation margin amounts are determined by relevant clearinghouses, although brokers might impose more stringent requirements on clients, and these margin requirements can vary as market volatility changes.

Whereas minimum margin requirements are often dictated by a relevant regulatory agency such as the Fed or CFTC, brokers, clearing firms, and clearinghouses are generally free to impose stricter requirements on their traders and members. Margin requirements will vary among brokers, instruments, etc., and will often be set based on the perceived risk of client accounts or clients themselves. Margin obligations for futures and options positions can be covered with cash and can sometimes be covered with other assets, contracts, or securities, such as government treasury instruments, bank guarantees, FX, CDs, and equities. Again, various clearinghouses and brokerage firms set their own rules for acceptance, subject to requirements imposed on them by higher authorities.

A popular risk and margin calculation system used and adapted for use by many clearinghouses and clearing firms is the *Standard Portfolio Analysis of Risk* (*SPAN*), created by the CME in 1988, particularly for use with futures and options portfolios. The older SPAN system was structured to analyze a grid or array comprised of 16 possible changes in market conditions within boundaries of the risk parameters set by the clearinghouse. The array was structured for use with an algorithm to calculate "scanning risk," which is considered to be the worst possible loss scenario for the portfolio. This analysis is then used to set the margin requirement for the portfolio. The OCC developed a somewhat similar method, known as the *Customer Portfolio Margin System* (CPM), for a risk assessment and margin calculation system used with options portfolios, based largely on its older Theoretical Intermarket Margining System (TIMS) structure. We will detail and discuss updates to the SPAN system later in Chapter 10.

In general terms, *hypothecation* is the pledge of assets as collateral for debt or other obligations, typically in scenarios involving mortgages, margin, short sales, and repo agreements. With respect to the hypothecation of securities, the asset owner typically does not give up ownership rights such as voting in corporate affairs and dividend or interest payments unless the owner defaults on her obligations. For example, should the broker issue a margin call that the client fails to meet, the pledged assets can be seized and sold.

Rehypothecation occurs when the securities broker re-uses assets pledged by a client as collateral as its own collateral; that is, the same asset is being used as collateral for both the client's loan and the broker's loan.

Delivery

Only a small fraction of listed futures and options contracts involve physical delivery of assets or even cash delivery at settlement or expiry, because most positions are closed out earlier. Thus, netting ensures that no further obligations are outstanding on closed positions. Futures and options exchanges and clearinghouses establish rules and manage the processes for physical delivery of assets underlying contracts or cash delivery. Of course, major U.S. derivatives exchange clearinghouses serve as central counterparties. Thus, on settlement and exercise dates, they make appropriate delivery to clearing members and take delivery of assets as per settlement or exercise obligations. Netting operations significantly simplify delivery procedures. Contract specifications set forth delivery obligations, including the settlement figure for futures contracts, sometimes referred to as the *Exchange Delivery Settlement Price* (EDSP).

When delivery procedures and policies are violated by clearing members, clearinghouses impose fines and other penalties, which can include trading suspensions and revocations of trading privileges, in addition to adverse publicity and other reputational costs for the trading member. Similarly, clearing and brokerage firms penalize their clients who fail to make appropriate deliveries.

3.5 OTHER FINANCIAL MARKET UTILITIES AND FUNCTIONS

The financial market infrastructure consists of numerous types of utilities, including clearinghouses and clearing firms. In addition, clearinghouses and clearing firms perform additional functions on behalf of their members and the corporations they serve. We discuss some of these services and utilities below.

Dividends

Stock value is primarily a function of dividends anticipated to be paid by the issuing firm, often on a quarterly basis. Ultimately, the prospect of payments made to shareholders of the firm are what attracts shareholder investment. A custody agent, often either a clearinghouse or a bank, facilitates this process for the dividend-paying corporation, collecting funds from the paying corporation and enabling the distribution of proceeds to shareholders of record. The process typically extends over several weeks, with four dates of particular importance. On an issuing firm's dividend *declaration date*, the company's board announces that a dividend will be paid. The stock trades *cum* (with the) *dividend* until the *ex-dividend date*, typically a few weeks later, at which time it begins to trade without the dividend (*ex-dividend*), which is set by the exchange to fall two days before the *date of record* announced by the issuing firm's board. The issuing firm actually makes the

payment on the *dividend payment date* as announced by the board through a paying agent or bank, which then transfers most of the payment to the DTC's affiliate Cede & Co. through Fedwire. Afterwards, the DTC distributes payments through its automated Dividend Service to appropriate members (brokerage firms, clearing firms, and banks), which then credit payments to their client accounts.

Stock dividends allocated and dividend reinvestment plan allocations requested by issuing firms are simply reflected in clearinghouse member accounts. Clearinghouse members then allocate these dividends to their customer (beneficial owner) accounts.

Transfer Agents

A *transfer agent* (*share registry* in the U.K., Australia, and New Zealand) is a third-party financial institution or individual that maintains records of ownership and transfers of shares and other securities issued by a publicly traded company.[5] Transfer agents might be trust companies, banks, or similar institutions that normally serve as paying agents of dividends, interest, and other disbursements to shareholders and other investors. Many transfer agents will handle company mailings to investors, deal with investor problems and issues, manage rights issues, reissue and cancel stock certificates (including certificates associated with stock splits), and tender offers, reorganizations, and bankruptcies as well as lost or stolen certificates. Many transfer agents will also manage dividend reinvestment plans (DRIPs), solicit proxies, and monitor corporate elections.

Debt Market Utilities

The fixed income discussions in Chapter 2 focused primarily on debt instruments and primary debt markets, largely involving the Treasury Automated Auction Processing System (TAAPS) and corporate instruments, whereas in this subsection, we focus more on secondary markets. Secondary OTC U.S. Treasury markets are enormous, including $14 trillion outstanding in roughly 390 different instruments in 2018, with an average daily trading volume of $530 billion.

U.S. regulations do not require Treasury security trades to be centrally cleared. In fact, most dealer-to-customer trades are normally cleared bilaterally, as are many dealer-to-dealer trades and trades by principal trading firms. Still, many dealer-to-dealer trades are executed through interdealer brokers and cleared centrally, particularly if both traders are members of the relevant central clearinghouse, which is normally the Fixed Income Clearing Corp. (FICC), a unit of the DTCC. As is the case with other central clearing markets, such trades are netted and involve novation by FICC. Members of these clearinghouses meet stringent risk-management standards, fulfill capital requirements, and contribute to default funds. The FICC has two primary divisions—the Government Securities Division (GSD), which clears Treasury and government agency issues, repos, and reverse repo issues (more than $1.4 quadrillion total cleared in 2019), and the

[5]See Securities and Exchange Commission (2015) for more details. Major transfer agents include Continental Stock Transfer & Trust, Computershare/BNY Mellon and American Stock Transfer & Trust.

Mortgage-Backed Securities Division (more than $67 trillion cleared in 2019). Clearstream clears and settles for a wide variety of European debt markets.

Settlement in U.S. Treasury markets is typically T+1 except for when-issued trades (trades executed before the instruments are issued) and a very small number of same-day cash settling trades. Treasury instrument trades can be executed by voice (phone) or electronically. Those executed by phone normally clear more slowly and are subject to more DKs (don't knows).

The NSCC clears and settles most U.S. transactions involving municipal and corporate debt instruments, whereas the DTC acts as the central securities depository for almost all municipal securities, even though this custody arrangement is not usually mandated. Corporate bonds are treated by the DTC in much the same underwriting manner as shares of stock. Interest payments are forwarded by issuing companies to paying agents or bond trustees, which pass those in the custody of Cede & Co. to the DTC; the DTC credits payments to member accounts (while forwarding payments to appropriate banks), which then allocate payments to client accounts (see Morris & Goldstein [2010]). The DTC also handles maturities of issues and call and conversion activations, transferring funds and making appropriate account changes as well as bankruptcy and reorganization announcements along with appropriate payments and member account adjustments.

3.6 SYSTEMICALLY IMPORTANT FINANCIAL MARKET UTILITIES

The nature of business for clearinghouses requires that they match long positions in securities and instruments against short positions in the same instruments. Every buy is matched with a sale. Exact matching provides a substantial protection against clearinghouse risk. Further protection is provided by clearinghouse margin requirements, clearinghouse capital, and a mutual default fund contributed to by clearinghouse members. However, what if a counterparty to a large transaction were to fail and be unable to fulfill its obligation to the clearinghouse? What if this obligation were long-term and allowed for substantial price volatility or movement before the settlement date of the instrument?

A central counterparty's *default waterfall* stipulates the sequence of resources available to the central counterparty or clearinghouse to protect the market and its participants from broker and clearing firm defaults.[6] If the defaulting trader's or client's margin is insufficient to enable its broker or clearing firm to cover the client's obligations, the broker or clearing firm then draws on the account balance of its client. If the full balance of the defaulting client's account is insufficient to cover the client's obligations, the broker or clearing firm is still required to make good to the clearinghouse on these obligations. The next protection in the default waterfall is the defaulting trader's broker or clearing firm margin with the clearinghouse, which can be seized to cover the default amount. If this member's margin proves insufficient to cover the obligation, the clearinghouse withdraws

[6]CPSS-IOSCO Principles (jointly issued by the Committee on Payment and Settlement Systems and the International Organization of Securities Commissions) provide for a set of minimum requirements for loss absorbency related to the default waterfall in order for central counterparties to maintain safety, stability, and transparency in the services they provide.

funding from the defaulting (brokerage firm or clearing firm) member's contributions to the clearinghouse's default fund. Should these funds prove insufficient, the clearinghouse draws on its own capital in the fund, and if necessary and permitted by the clearinghouse rules, draws on the nondefaulting clearinghouse members' contributions to the clearinghouse's default fund as well. Should all of these resources prove insufficient to cover default obligations, each clearinghouse will maintain a sequence of recovery tools, which might include assessments or additional contributions of assets by clearinghouse members, which stipulate how its members will cover all obligations.

Fortunately, central counterparty failures are rare; even clearing member defaults are unusual (see McPartland and Lewis [2017]). A major exception was the 1974 failure of the French Caisse de Liquidation, the result of unmet margin calls by commodity dealers and the clearinghouse's failure to increase margin requirements. The Malaysian Kuala Lumpur Commodity Clearing House failed in 1983 for similar reasons. There have been several near failures, including the Hong Kong Futures Exchange, which was closed for 4 days in 1987 and required a government influx of cash. As we will discuss later, sometimes governments provide for bailouts of systemically important financial market utilities. All three of these central counterparties cleared long-term derivative contracts, which exposed them to additional risks. Both the Chicago Mercantile Exchange and the Options Clearing Corporation experienced stress events due to unmet margin calls resulting from the 1987 stock market crash. Neither institution defaulted, but investors were concerned about their viability. The commodities default fund for the Swedish Nasdaq Clearing AB in 2018 was more than 50% depleted by the failure of a single trader to meet a margin call. Nevertheless, despite these exceptions over the years, clearinghouse and central counterparty arrangements have generally tended to withstand even severe crises.

Financial market utilities are institutions that provide the essential infrastructure for transferring, clearing, and settling financial transactions. Many of these utilities also serve as central counterparties. *Systemically important financial market utilities* (SIMFUs) are U.S. financial market utilities that have been designated by the U.S. Financial Stability Oversight Council (see Chapter 5) as being among those whose material financial distress or failure could pose a threat to the financial stability of the U.S. Title VIII of the Dodd-Frank Act of 2012 provided for the designation of both the Financial Stability Oversight Council and SIMFUs, and for potential access to emergency liquidity at the discretion of the Federal Reserve System.

One important example of a SIMFU is the *Clearing House Interbank Payments System* (CHIPS), the largest private sector (it is owned by its member banks) clearing system in the world, clearing and settling more than $1.8 trillion daily. The failure or even disruption of CHIPS activity would have a devastating impact on the functioning of the U.S. financial system. The following institutions have received the SIMFU designation due to their critical roles in U.S. financial markets infrastructure:

The Clearing House Interbank Payment System (CHIPS)
CLS Bank International (provides settlement services in FX markets)
The Chicago Mercantile Exchange, Inc. (CME)
ICE Clear Credit, LLC (clears credit default swaps)
The Options Clearing Corporation (OCC – listed options and certain futures)

The Depository Trust Corporation (DTC, a unit of DTCC)
Fixed Income Clearing Corporation (a unit of DTCC)
The National Securities Clearing Corporation (NSCC, a unit of DTCC)

These SIMFUs face enhanced regulatory oversight by the Federal Reserve System, the Securities Exchange Commission, and the Commodities Futures Trading Commission. For example, the Division of Clearing and Risk of the CFTC oversees derivatives clearing organizations (DCOs), the clearing of swaps, futures, and options on futures, and market participants that may pose risk to the clearing process. Furthermore, SIMFUs face heightened standards for their boards of directors and heightened expectations for comprehensive risk management systems and liquidity maintenance along with detailed and viable plans for recovery and resolution.

First, although rare, clearinghouses do fail. Should a SIMFU experience a stress event, a first step to recovery is normally to draw from the accounts of and to demand additional margin from members or specific members associated with the stress. The second line of defense might be the SIMFU's own capital, followed by injections from the SIMFU's mutual default fund, if such a fund is maintained. Most clearinghouses then provide for their members to inject additional capital as needed. Without detailed recovery and resolution plans, additional lines of defense or resolution, including potential government bailouts, tend to be rather ambiguous, even in the U.S.

Additional Reading

Unpublished papers prepared by the Securities and Exchange Commission (2020) and Jones (2018) provide a substantial amount of detailed information on SIPs and the CTA along with the reporting data that they offer traders. Jones (2018) provides a wide-ranging discussion of securities market data itself as an essential product, sold by exchanges and other venues to financial institutions, media, and individuals in competitive markets. Costs and benefits of information to traders are discussed as well. Loader (2020) is an easy-to-follow and fairly comprehensive book providing useful discussions on securities clearing and settlement along with various clearinghouses and related institutions. Morris and Goldstein (2010) do not discuss clearing and settlement with as much detail as Loader, but their monograph offers an even more basic introduction to IPOs and other market issues. Their earlier book (Morris & Goldstein, 2009) also provides a readable, basic introduction to clearing and settlement. A more technical discussion of clearing, margin calculations, and central counterparties, particularly in derivatives markets, is provided by Gregory (2014).

References

Bernstein, A., Hughson, E., & Weidenmier, B. (2019). Counterparty risk and the establishment of the New York Stock Exchange Clearinghouse. *Journal of Political Economy*, *127*(2), 689–729.

Depository Trust & Clearing Corporation. (2020). Settlement. Web user guide. Unpublished mimeo. February 21, 2020. Available at <www.dtcc.com/products/training/helpfiles/settlement_help/help/settlement_web.pdf> Accessed 18.09.21.

Faruqui, U., Huang, W., & Takáts, E. (2018). Clearing risks in OTC derivatives markets: The CCP-bank nexus. *BIS Quarterly Review*, December, 73–90.

Gregory, J. (2014). *Central counterparties: Mandatory clearing and bilateral margin requirements for OTC derivatives.* Chichester, United Kingdom: Wiley.

Jones, C. M. (2018). Understanding the market for U.S. equity market data. Available at <www.sec.gov/comments/4-729/4729-4545881-176154.pdf> Accessed 18.08.21.

Loader, D. (2020). *Clearing, settlement and custody* (3rd ed.). New York: Butterworth-Heinemann.
McPartland, J. W., & Lewis, R. (2017). The Goldilocks problem: How to get incentives and default waterfalls "just right". *Federal Reserve Bank of Chicago: Economic Perspectives, 41*(1).
Morris, V. B., & Goldstein, S. Z. (2009). *Guide to clearance & settlement*. New York: Lightbulb Press.
Morris, V. B., & Goldstein, S. Z. (2010). *Life cycle of a security*. New York: Lightbulb Press.
Securities and Exchange Commission. (2015). Advance notice of proposed rulemaking, concept release, and request for comment on transfer agent regulations, Release No. 34-76743. Unpublished mimeo. Available at <www.sec.gov/rules/concept/2015/34-76743.pdf>
Securities and Exchange Commission. (2020). Data market infrastructure, Release No. 34-90610. Unpublished mimeo. Available at <www.sec.gov/rules/final/2020/34-90610.pdf>
SIPC (2021). History and track record. Available at <https://www.sipc.org/about-sipc/history> Accessed 13/02/2022.

3.7 EXERCISES

1. Some stock brokerage firms have offered temporary "specials" where they don't earn commissions on stock trades and a few (e.g., Robinhood Financial, discussed in the chapter) have made zero stock trading commissions permanent part of their business model. In fact, most stock brokerage firms earn less than 40% of their revenue from commissions. How can such firms realize revenues from their clients or elsewhere if they don't produce stock-trading commissions?
2. Congress and the SEC have attempted to enhance market competition among U.S. trading venues without intensifying the harmful effects of market fragmentation. What is market fragmentation, and what are its effects that might adversely affect securities markets? How might SIPs mitigate these adverse effects caused by multiple market venues?
3. The SIP evolved from the ticker tapes providing price and volume data from the exchanges. It is said that now, the SIP is far busier than it was more than a century ago when the ticker tapes were significant. Why is this?
4. Each stock trading venue provides quote and trade information to the SIP, which disseminates the information to the market. On which tape would the NBBO for shares of the Bats-traded Global X U.S. Infrastructure Development ETF (PAVE) be reported? Why?
5. Describe how real-time clearing differs from batch processing. Why is netting less expensive than real-time processing?
6. *Netting* and *novation* are important activities of most clearing firms. What is the difference between these two activities?
7. Support the argument that central counterparties (novation) in derivatives markets standardize credit risk for all traders.
8. The table below represents transactions executed during a given trading day in a given stock market:

Trader	Broker	Stock	Buy/Sell Shares	Cash Amount	Settling Bank
A	Merrill	General	Buy 100	$5,000	City
A	Morgan	General	Sell 200	$10,000	Chase
B	F-Trade	DotCom	Buy 100	$6,000	BoB

(*Continued*)

C	F-Trade	R-Train	Buy 100	$8,000	BoB
C	Schwalb	DotCom	Sell 100	$6,000	Chase
D	Schwalb	R-Train	Sell 100	$8,000	Chase
D	Schwalb	General	Sell 300	$15,000	Chase
D	Schwalb	General	Buy 400	$20,000	Chase

a. What is the net change of shareholdings in each of the broker accounts with DTC?
b. What is the net change of shareholdings in each of the client accounts with each broker?
c. What is the net change of the DTC's cash account with each settling bank?
d. What is the net-net cash amount due or paid to the DTC?

9. Some observers suggest that real-time clearing (T + 0) would improve market efficiency relative to (T + 1) or (T + 2) clearing.
a. How would real-time clearing affect the netting process?
b. How would real-time failure affect the number of out-trades or DKs (don't knows) in the clearing process?

10. During the winter of 2021, Robinhood Securities, a brokerage firm popular with retail investors, temporarily restricted trading certain highly volatile securities, including shares of GameStop. The DTCC had become concerned that Robinhood would be unable to fulfill its rapidly increasing collateral requirements. Why would Robinhood risk alienating its customers by restricting trading in GameStop and other securities experiencing such frenzied trading? What else might Robinhood have done to assure the DTCC that it would be able to fulfill its collateral requirements?

11. Long-term U.S. Treasury bonds are considered to be very high-quality investments with virtually zero default risk. Is it reasonable to assume that the risk of settling U.S. Treasury bonds is also of very low risk to the clearinghouse?

CHAPTER

4

Institutional Trading

4.1 INSTITUTIONS AND MARKET IMPACT

Institutions today own the bulk of securities traded in today's marketplace and execute the majority of trades. The Boston Consulting Group (2021) estimated that institutions managed roughly $103 trillion in 2021 for tens of millions of clients. Among these institutional investors are buy-side institutions such as mutual funds, pension funds, life insurance companies, trust departments of banks, and investment companies. Institutional investor transactions are frequently executed by professional traders, either as employees of the institutions or acting as their agents, and these transactions provide opportunities and risks for other traders. Institutional transactions also have important effects on security liquidity, prices, and volatility. Institutional investors play crucial and varied roles in markets and are behind conditions that affect every trader. For example, stocks with the largest proportional institutional ownership experience the most volatile swings when the market is most volatile.[1] Why is this? Is it because institutions invest in larger blocks of shares and their transactions push the market more than noninstitutional shareholders? Are institutional shareholders more informed than other traders? Regardless, understanding the link between institutional trading and security behavior is important to the trader. We will discuss such issues in this and in later chapters.

We introduced buy and sell sides of the market in Chapter 1. In this and the next two sections, we will discuss buy-side institutions before turning our attention to sell-side institutions. On the buy side, we start with the *investment company*, which is an institution that accepts funds from investors for the purpose of investing on their behalf. The investment company is compensated with fees, typically based on some combination of performance and investment portfolio size. Clients expect that the investment company will provide professional and competent management services and pass on cost advantages resulting from economies of scale. These cost advantages include reduced transactions costs and managerial costs. In addition, these scale advantages provide for more efficient record keeping and enable improved diversification. There are a number of different types of investment companies that register with the Securities and Exchange Commission (SEC)

[1]See Gompers and Metrick (2001) and Dennis and Strickland (2002).

DOI: https://doi.org/10.1016/B978-0-323-90955-6.00004-5

© 2023 Elsevier Inc. All rights reserved.

and operate under SEC oversight. In addition, there are a variety of unregistered investment companies. We will discuss many of these shortly.

As we mentioned above, institutional traders can have dramatic impacts on security prices. Obviously, large block orders will result in dramatic price reactions. But how large will these reactions be, and what other factors will affect these reactions? Which types of institutional transactions will have the most significant price impacts? For example, numerous studies, including Kraus and Stoll (1972) and Saar (2001) have documented asymmetric reaction to block trades. In particular, institutional block buy orders result in a much larger upward effect than the downward effect caused by otherwise comparable institutional sell orders. This raises an interesting question: Why are institutional buys associated with significantly larger price impacts than institutional sells? Arguments offered by Chan and Lakonishok (1993) and Saar (2001) suggest that buy orders convey more information than sell orders because:

1. Liquidity needs drive many sales.
2. Most institutions have a much larger pool of stocks from which to buy than to sell.
3. Mutual funds are reluctant to short sell, so that they focus their research efforts on shares that they might buy and might serve as good diversification mechanisms.
4. Mutual funds cannot borrow to invest, so that they have to be particularly selective in their purchases.

Thus, if one or more of these observations are correct, there should be more information content in institutional stock purchases than stock sells, and price changes will react accordingly to institutional transactions. We will focus more on the market impact of institutional transactions later in this chapter and in Chapter 6.

4.2 REGISTERED INVESTMENT COMPANIES

The buy-side institutions that execute the largest volume of transactions have an enormous impact on trading for all market participants. Among the most significant of these managed investment institutions are the registered investment companies, which are required to register with the Securities and Exchange Commission (SEC) under the *Investment Company Act of 1940*. Such companies include mutual funds, unit investment trusts, and real estate investment trusts (REITs). Investment companies manage assets on behalf of clients. Managed assets usually include, but are not limited to, publicly traded financial securities such as shares of stock, bonds, and derivative instruments.

Managed Investment Company Types

One type of managed investment company is the *closed-end investment company*, which is a corporation that issues a specified number of shares that can be traded in secondary markets such as the New York Stock Exchange. One might purchase shares through a brokerage firm. *Open-end investment companies*, typically known as *mutual funds*, frequently accept additional funds directly from investors and are willing to repurchase or redeem

outstanding shares. An investor buys into a mutual fund by purchasing shares from the fund and cashes out by redeeming those shares.

The major markets in which mutual funds operate are equities, fixed income, money markets and commodities and foreign exchange markets. Fixed income funds seek to provide stable interest income for their investors. One type of income fund, the high-yield fund, seeks to provide higher income by investing in a diverse portfolio of higher-risk, higher-return (junk) bonds. Municipal bond funds purchase tax-free state and local government bonds for their investors. These funds are particularly attractive to investors in high income-tax brackets. Money market funds seek to maintain safety and liquidity for their investors, focusing their investments in short-term, highly liquid, and safe debt instruments. Such investments include U.S. Treasury bills and bonds, GNMA, FNMA, and other federal government agency issues, bank issues such as repurchase agreements and corporate issues such as commercial paper. Stock funds might be categorized according to their objectives or types of stock that they may invest in. Among the categories of stock funds are aggressive growth funds, growth funds, income funds, balanced funds, global funds, and emerging markets funds. Some larger fund management companies such as Fidelity, Dreyfus, Vanguard, and T. Rowe Price manage entire families of funds with a wide variety of investment objectives ranging from aggressive growth to generating tax-free income. Commodity funds enable investors to take focused or diversified positions in a variety of commodities such as oil, gold, and agricultural products, and foreign exchange funds enable investors to take positions in the currencies and instruments of other countries.

Index funds such as those in the Vanguard family have increased in popularity in recent years. These funds employ a passive management technique where managers simply attempt to maintain a fund composition matching a particular market index (basket of stocks or other instruments) rather than attempt to "out-guess" the market. The passive management technique is also intended to keep fund management expenses low because the index strategy does not require hiring securities analysts to pick stocks in an effort to outperform the market. In addition, these funds provide investors with opportunities to achieve broad diversification at a low cost. As of year-end 2020, and excluding ETFs, the U.S. index fund market was comprised of 490 funds with $4.807 trillion in assets (Investment Company Institute, 2021).

Managed Investment Company Fees and Structures

Mutual funds that accept investments directly from investors without a sales charge are called *no-load* funds; investors buy and sell funds at *net asset value*, the difference between the fund's total assets and liabilities divided by its number of outstanding shares. An investor buys into a mutual fund by paying for her shares exactly the value of her proportional ownership of the fund's investments, computed at the end of the transaction date, without a sales fee or commission. Institutions that charge a sales fee are called *load* funds. The loads may be imposed when investors buy into a fund (*front-end load*), sell out of a fund (*back-end load* or redemption fee), or a combination of both. An investor should be aware that fund performance computations are frequently overstated because returns are usually calculated only on the investment net of fees. In addition, load charge percentages

are generally understated in that the loads are determined as a fraction of the amount invested in the fund in addition to any loads charged.

A number of funds adopt a 12b-1 plan that enables them to use fund assets to market their shares. Such 12b-1 expenses are often included with the sum of administrative expenses when computing annual expense ratios. Funds are permitted by the SEC to call themselves no-load funds even if they charge an annual 12b-1 fee, as long as the amount is less than 0.25% of the invested amount per year. The 12b-1 fee in effect enables the fund to impose its marketing costs on investors without reporting them as loads. In addition to potential load and 12b-1 fees, both load and no-load funds normally impose an annual management fee. These fees cover management compensation and various periodic administrative expenses.

A managed investment company such as a mutual fund maintains a board of directors to oversee its operations, but normally does not have managers and employees, at least not in significant numbers. Most managed investment companies retain a separate *management company* or *advisor* to actually manage the fund's assets in accordance with the fund's prospectus. The advisor will typically hire analysts and fund managers and will pay the following types of institutions to perform certain important functions:

Custodian: Holds the fund assets on behalf of shareholders.
Transfer agent: Processes orders to purchase and redeem shares of the fund and maintains customer records. The NSCC (see Chapter 3) processes and clears most mutual funds that involve brokers and dealers.
Distributor: Markets shares of the fund through various channels.

Shareholders of management companies incur management fees, miscellaneous administrative expenses, and often other costs as well.

ETFs

Exchange-traded funds (*ETFs*) are funds whose shares trade on exchanges. The first was the *S&P Depository Receipt* (*SPDR* or "Spider," ticker SPY) sponsored by State Street Bank and the Merrill Lynch Investment Company. The Spider fund is intended to mimic performance of the S&P 500 Index by maintaining the same portfolio of securities as is comprised by the index. As of year-end 2020, the U.S. ETF market was comprised of 2204 funds with $5.4 trillion in assets, accounting for approximately 69% of the $7.9 trillion worldwide ETF market (Investment Company Institute, 2021).

Unlike the case with most mutual funds, investors can trade shares of ETFs throughout the day at market prices that vary as the market index varies. Hence, investors can more easily trade the market portfolio throughout the day rather than have to wait for closing net asset values (NAVs) as is required for trading mutual funds. This ease of trading facilitates creation of options and other derivative contracts on the indices, which do not normally exist on mutual funds. Additionally, because an ETF is not actively managed, investors can benefit from low management expenses. On the other hand, investors typically will have to pay brokerage expenses to trade the fund and will be faced with a bid−offer spread, potentially increasing transaction costs. Some ETFs are leveraged and

investors might have the opportunity to purchase ETFs on margin or even to short sell them. Most mutual funds cannot provide these opportunities. Discrepancies between ETF market prices and those of their index target portfolios create potential arbitrage opportunities for ETF market participants.

An ETF is created by a *sponsor* who first chooses the ETF's investment objectives and/or index target along with the method of tracking the index. The sponsor will form a trust or other entity as the holding vehicle for securities that the fund will hold. The sponsor then identifies a portfolio of assets (e.g., a portfolio that will replicate the index target) in which to invest. This portfolio will typically comprise some combination of shares of equity, bonds and/or total return swap contracts.[2] The sponsor will then select an institutional investor to serve as the fund's *authorized participant*, which, in turn, purchases the actual portfolio of assets to form *creation units* that will correspond to a fixed number of shares in the ETF. Thus, the authorized participant (there might be more than one) actually purchases and delivers the portfolio of assets that the ETF will hold (initially called the creation units or creation basket), in exchange for ETF's shares that are taken back from the fund's sponsor who created them. Then all or most of the ETF shares are listed and sold to other investors on an exchange. Under normal circumstances, no new shares are created and no shares are redeemed during the life of the fund, enabling the ETF to closely match its index target and to avoid holding cash. Exceptions might occur when actual demand for fund shares differs substantially from the expected demand. When the ETF is ultimately terminated, the authorized participant repurchases shares from investors and redeems them to the fund sponsor.

For example, the trust underlying SPDR contracts is comprised of a portfolio of individual stocks that replicates the S&P 500. On the other hand, bear funds such as the *Short S&P 500 ProShares* (ticker SH), created to enable investors to easily short the market index, are comprised of a combination of short positions in shares of stock and short positions in swap contracts involving shares. *Ultra ETFs* (bull funds) are created to enable investors to easily leverage their investments in an index. For example, the *Ultra S&P 500 ProShares* (ticker SSO) invests in combinations of shares of stock along with long positions in shares in swap contracts to leverage its position in the market index. Funds that double leverage (2x) or triple leverage (3x) are most common among bull funds, but funds are available with as much as 50x leverage.

Other well-known ETFs include *DIA* "Diamonds" that mimic the Dow Jones Industrial Average, *QQQ* that mimics that NASDAQ 100, and a number of narrower sector and industry funds such as the *i-shares energy sector* that mimics the portfolio of Dow Jones Energy companies. These funds all provide investors opportunities to trade large or important small segments of the market and to hedge risks associated with these segments.

While our ETF discussion has thus far focused on passively managed funds that seek to replicate index targets, at year-end 2020, there were 2204 actively managed ETFs with almost

[2]Total return swaps normally provide for one party to swap cash flows associated with a particular basket of securities for cash flows associated with an interest rate instrument. Thus, for example, a swap can provide for the exchange of profits or losses on a stock market index for cash flows realized from a treasury instrument. Sometimes, funds will substitute options, forward, futures, and reverse repo agreements for swap contracts to create leverage.

$174 billion under management in the U.S. (Investment Company Institute, 2021). These funds seek to fulfill some investment objective other than index replication. Shares of these funds trade as shares of individual stock, but transparency of fund management activities is essential because fund shares are traded in secondary markets throughout the trading day.

While different in structure from ETFs and index funds, unit investment trusts have been created to replicate indices. For example, *Diamonds* are shares of the Diamonds Trust created to replicate the Dow Jones Industrial Average (DJIA) and to sell on the former American Stock Exchange (ASE).[3] *Spiders* (Standard & Poors Depository Receipts) track the S&P 500, *Midcap Spiders* track the S&P Midcap 400 Index, and *Webs*, which track the Morgan Stanley world equity benchmark indices, were also created to trade on the ASE. Combined, these four products accounted for approximately 15% of ASE volume prior to its acquisition by the NYSE. The "Dogs of the Dow" track the 10 DJIA stocks with the highest dividend yields, and are represented by their own trust. Although each of these trusts are designed to track a particular index, their shares are often more volatile than the index itself. One explanation for this volatility is that there is substantial trading volume when the market swings. This possible explanation is interesting in that it might suggest that traders' activities tend to increase market volatility.

4.3 UNREGISTERED INVESTMENT COMPANIES

As we discussed, registered asset managers such as mutual funds provide investment services for tens of millions of clients, functioning under the regulatory authority of the SEC. But there are many asset managers that are not required to register as registered investment companies. For example, pension funds and private equity management firms are not required to register with the SEC as public investment companies. Pension fund clients cannot trade their fund investments. *Private equity* firms are alternative professionally managed equity investments in unregistered securities issued by public and private firms. Investments in private equity firms are often tightly restricted by their managers, in part, to avoid SEC reporting requirements for registered investment companies. Other unregistered asset managers include banks, bank trusts, foundations, endowments, and insurance companies. All of these institutions are important buy-side securities traders.

Pension Funds

Pension funds are established by employers to facilitate and organize the investment of employees' retirement funds. *Defined benefit plans* specify payments that employees will receive when they retire and *defined contribution plans* define employer and employee contributions, but actual benefits depend on fund investment performance. In most cases, investment horizons for pension funds are long term. Many pension funds are large enough to afford more sophisticated investment techniques and managers. Pension funds in sum hold over $20 trillion in assets for their beneficiaries. More than half of working Americans participate in pension plans, including defined contribution plans such as 401-k

[3]The ASE has since been integrated into the NYSE.

plans, representing very diverse ownership structure somewhat representative of the U.S. population. Regulators, in part due to the Employee Retirement and Security Act of 1974 (ERISA), tend to discourage pension plans from taking imprudent risks. Nevertheless, many pension funds, particularly defined benefit plans, have been criticized for being underfunded and not providing adequate capital to secure their employees' retirements (see, for example, Kwan, 2003). The Pension Benefit Guarantee Corporation, a federal corporation created by the ERISA, provides limited protection for the pensions of nearly 44 million American workers and retirees in over 30,000 private, single-employer and multiemployer, defined-benefit pension plans.

Bank Trusts and Private Banking

Banks also provide professional asset management services, including trust management for clients. Trusts are legal entities or vehicles for enabling settlors or grantors (entities that set aside assets in trusts on behalf of beneficiaries such as heirs, charities, and others) to accomplish various financial goals. Banks in their roles as trustees or asset managers actively manage trusts on behalf of grantors and their beneficiaries. Most banks maintain trust departments to serve as trustees or perform services on behalf of trustees.

Private banking refers to financial services such as banking and investment management services provided by banks to private high-net-worth clients. Private banking services offered by major banks such as Citigroup, HSBC, Bank of America (U.S. Trust), and Wells Fargo have frequently been particularly lucrative. These personalized services tend to be rather broad-based and banks strive to provide their clients the greatest confidentiality. Private banking services frequently include specialized financing solutions to complex problems, as well as retirement and estate planning. These operations contribute to banks' trading operations.

Private Equity

Private equity refers to asset managers who make equity investments in companies that are not publicly traded. Private equity markets provide funding for start-up firms, private middle-market firms ($25 million to $500 million in sales), management buyouts (MBOs), leveraged buyouts (LBOs), firms in financial distress, and public firms seeking buyout financing. Traditionally, private equity ownership is not publicly traded (there are a few exceptions), not liquid, and exempt from SEC registration requirements. Private equity firms frequently take active monitoring, management, and advising roles in their portfolio firms. There are a variety of types of private equity firms, ranging from venture capital (VC) firms to hedge funds.

Hedge Funds

Hedge funds are unregistered private funds that allow investors to pool their investment assets. To avoid SEC registration and regulations, hedge funds usually only accept funds from small numbers (often less than 100) of *accredited* investors, typically high-net-worth

individuals and institutions. Because most hedge funds have only a small number of managers, they typically focus their investment strategies on the expertise of a smaller number of key managers. Many hedge funds seek investment opportunities or niches where larger institutions are prohibited or constrained due to regulatory restrictions. For example, because many banks, pension funds, and other institutions cannot focus activities in the securities of distressed corporations facing reorganization or bankruptcy, some hedge funds will specialize in such investments. Other funds may specialize in short sales and derivatives to hedge against market downturns and others will simply focus on searches for arbitrage opportunities. Hedge fund managers typically take a proportion of assets invested (2% is a norm) and another portion of profits (20% is typical) as compensation. While hedge funds are frequently able to report results that beat the market, investors should realize that performance results do not usually include the last several months of a hedge fund's existence, which is when a fund is most likely to fail (pretty much by definition). The point here is that while the risky strategies employed by many hedge fund managers can consistently produce strong returns over long periods, many lose large sums and close down in remarkably short periods following long periods of success. Sixty consecutive months of market-beating returns followed by a single month with a return of −100% is obviously a disaster for investors. For example, consider the long-running successes of Long Term Capital Management and Amaranth, after which both suffered spectacular meltdowns (see Hosking, 2006).

4.4 BEST EXECUTION, EXECUTION COSTS, AND PRICE IMPROVEMENT

"*Best execution* refers to traders receiving the most favorable terms available for their trades" (Macey and O'Hara, 1997). Brokers have a responsibility to provide their clients best execution, though, as we will discuss in the next chapter, this obligation has largely been replaced by best price. Unfortunately, because many factors are needed to define the most favorable terms, best execution is not an easy concept to quantify, evaluate, or enforce. Suppose, for example, that the best quotes are available in one market, but trades are more likely to execute between quotes in another market. Which market will provide for the best execution? What if one market normally executes at a better price than another market, but takes longer with more risk? Regardless, execution costs are an important factor in evaluating best execution. Execution costs might be considered to be the sum of (1) order processing costs and (2) market impact costs.

Measuring *order processing costs* is normally fairly straightforward. These costs will normally include brokerage fees, which might account for other processing costs as brokers normally provide these services as well. However, *market impact* or slippage costs (introduced in Sections 1.6 and 2.5) measurement is more problematic. The price or market impact of a transaction, the effect that a given transaction has on the market price of a security, is particularly important for a large transaction or in an illiquid market. When institutions execute large transactions, their buy or sell activities affect security prices, typically in a manner that increases execution costs. While substantial, especially for large orders, these costs are not explicit, and might be considered hidden.

Spreads

Investors, particularly more active institutional investment managers need to track execution quality to measure the trading performance of their investment brokers. Brokerage firms have traditionally used bid−offer spreads as a basis for charging commissions to their customers. That is, buyers frequently bought at the offer and sellers sold at the bid, with brokers pocketing the difference (spread) as their commission. This tradition is one reason that spreads are an important factor for gauging transaction costs.

While this spread-based commission model has changed somewhat in many markets, many studies concerned with execution costs have used the *quoted spread* (offer minus bid) or *half-spread* (quoted spread/2) to measure execution costs. However, when transactions are executed inside spreads, these measures can overstate execution costs (Macey and O'Hara, 1997). In this case, the *effective half-spread*, the absolute value of the difference between the actual transaction price and the mid-point of the spread may be a more accurate measure of the cost of the transaction. However, none of these measures are likely to capture the *market impact* of a transaction or series of transactions, often referred to as *slippage*. That is, each transaction is likely to force the price in some direction, upwards for purchases and downwards for sales. This is the market impact of the transaction. The *realized half-spread*, also called the *perfect foresight half-spread*, is the absolute value of the difference between some future price and the trade price. This realized half-spread might be particularly relevant when the transaction was executed as part of a series of transactions. Here, the assumption is that the subsequent or future price is the fundamental value of the security and its difference from the prior price reflects the price impact of the transaction or series of transactions. The "look-ahead window" for the subsequent price might be defined as one execution subsequent to the relevant transaction, some other number of executions subsequent, one or more of a different investor's executions subsequent to the relevant execution, or a given length of time subsequent to the relevant execution. Presumably, the subsequent price reflects the "intrinsic value of the security" whereas the relevant execution occurs at a price that reflects the slippage caused by the relevant investor's order. Again, this perfect foresight half-spread is particularly relevant to the trader if the trade was part of a series of trades.

Suppose, for example, that your stock purchase transaction was executed at 5.03 when it was quoted at 5 bid and 5.05 ask, and that someone else's subsequent transaction executed at 5.04. The quoted spread was $5.05 - 5 = 0.05$, the half spread was $(5.05 - 5)/2 = 0.025$, and the effective half-spread was $|(5.03 - 5.025)| = 0.005$. The realized half-spread, based on the immediately subsequent transaction, $|5.04 - 5.025|$ was 0.015. For research and analysis purposes, sometimes spreads are computed as proportions of transaction prices or mid-points.

Since spreads represent potential sources of income to brokers and market-makers, it is important to know what their determinants are. Among the factors that affect spreads are liquidity, volume, volatility, skewness, and information asymmetry. These factors and their effects on spreads are discussed with more detail in Chapter 6.

Some efforts at measuring market impact have focused on examining prices of consecutive transactions. For example, what if we want to see how a series of our 80 buy orders for a stock in one day have impacted the price of that security? Presumably, each of our buy executions moved or could have moved the security price. An analysis of this series

of trades with associated volumes might indicate the relationship between order sizes and market impact. We might be able to use a regression analysis of price changes on transaction sizes to infer a relationship between security price movements and transaction sizes. We might be able to improve this understanding by adding appropriate control variables to the regression analysis. We will discuss such relationships in the next section.

Numerous academic and regulatory studies have compared execution costs of different markets. Many older studies (e.g., Huang and Stoll, 1996) found that execution costs on NASDAQ were substantially higher than on the NYSE and that other equity markets did not compare favorably to the NYSE. Christie and Huang (1994) found that NYSE effective half-spreads averaged 7.9 cents compared to 18.7 cents on NASDAQ and that trading costs fall for firms that switch listings from the NASDAQ to either the AMEX or the NYSE. Similarly, Huang and Stoll (1995) found that 26.7% of trades on the NASDAQ were inside the spread, compared to 37.9% of trades on the NYSE. Thus, NYSE transactions were more likely to exhibit price improvement.

Price improvement occurs when an order is executed a price better than the prevailing bid or offer. Exchange specialists and market makers had or have an exchange responsibility to seek to improve markets for securities that they trade. Price improvement typically occurs when a specialist or market maker allows market orders to execute inside the spread, either by improving their quotes or by allowing market orders to execute against other submitted market orders inside the spread.

Illustration: Slippage and the Perfect Foresight Half Spread

Table 4.1 depicts a series of 20 historical transaction prices that a hypothetical brokerage firm paid for a given stock's purchases. Table 4.1 also depicts subsequent transaction prices, from which perfect foresight half-spreads are computed. The sizes of the purchases follow, so that the broker can analyze the relationship between the perfect foresight half-spread and the transaction size based on 20 of the broker's own purchase transactions for the same stock. For example, the first transaction has the broker purchasing 1000 shares for $50.00. The next transaction, by another investor, was at a price of $50.30, implying a perfect foresight half-spread of 0.30. Purchases by other investors are omitted in the "Purchase Price" column. Note that larger purchase orders tend to lead to higher perfect foresight half-spreads in Table 4.1. This suggests that larger orders lead to greater slippage, just as we would expect. What might be less apparent in Table 4.1 is that the proportional increase in slippage increases as the proportional size of the transaction increases. That is, the relationship between slippage and transaction size is concave up.

We can analyze the relationship between the perfect foresight half-spread and transaction sizes with an ordinary least square (OLS) regression. Suppose the broker expects that this relationship will not be linear because the effect of transaction size on the perfect foresight half-spread is expected to be concave up. That is, for example, the broker might expect that a single purchase transaction for 1000 shares will cause more slippage than 10 purchase transactions for 100 shares each. To analyze this nonlinear relationship, we will conduct a log-log regression of the following form:

$$log(Perfect\ Foresight\ Half\text{-}Spread)_t = logb_0 + b_1 log(Transaction\ Size) + \varepsilon_t$$

TABLE 4.1 Transaction Size and Slippage

Purchase	Purchase Price	Subsequent Transaction Price	Perfect Foresight Half-Spread	Transaction Size
1	50.00	50.30	0.30	1000
2	50.26	50.30	0.04	200
3	50.26	50.35	0.09	400
4	50.17	50.33	0.16	600
5	50.32	50.33	0.01	100
6	50.41	50.46	0.05	300
7	50.51	50.55	0.04	200
8	50.44	50.47	0.03	300
9	50.71	51.12	0.41	1200
10	51.03	51.04	0.01	100
11	50.97	50.99	0.02	200
12	50.57	50.75	0.18	700
13	50.71	50.74	0.03	200
14	50.92	50.96	0.04	200
15	50.91	50.93	0.02	200
16	51.18	51.32	0.14	600
17	51.71	51.76	0.05	300
18	52.04	52.12	0.08	400
19	52.03	52.34	0.31	1000
20	52.30	52.34	0.04	300

This implies the following estimated relationship between the perfect foresight half-spread and transaction size:

$$Perfect\ Foresight\ Half-Spread = v(Transaction\ Size)^m$$

where b_0 serves as an estimate for v and b_1 serves as an estimate for m. These parameters b_0 and m characterize slippage, which is positive when v is positive and m exceeds 1. Higher values for b_0 and m indicate increased slippage.

Table 4.2 presents *log* values (base 10) for the perfect foresight half-spreads and transaction sizes, and depicts regression results. The log-log regression implies that $b_0 = -4.96$ and $b_1 = 1.48$; that is, v is approximately equal to 0.00001 and m is approximately equal to 1.5 (use anti-logs of b_0 and b_1 to obtain v and m). Thus, slippage, $S = vc^m \approx 0.00001c^{1.5}$ where c is transaction size. Regression t-statistics, the F-statistic, and the adjusted r-square value all suggest that these results are statistically significant.

TABLE 4.2 Slippage and Transaction Size Regression

Log Perfect Foresight Half-Spread	Log Transaction Size						
−0.522879	3.00000	SUMMARY OUTPUT					
−1.397940	2.30103						
−1.045757	2.60206	Regression Statistics					
−0.795880	2.77815	Multiple *R*		0.977851577			
−2.000000	2.00000	*R* Square		0.956193706			
−1.301030	2.47712	Adjusted *R* Square		0.953760023			
−1.397940	2.30103	Standard Error		0.101914919			
−1.522879	2.47712	Observations		20			
−0.387216	3.07918						
−2.000000	2.00000	ANOVA					
−1.698970	2.30103		df	SS	MS	*F*	Significance *F*
−0.744727	2.84510	Regression	1	4.0809	4.0809	392.9	1.12419E − 13
−1.522879	2.30103	Residual	18	0.18696	0.010386651		
−1.397940	2.30103	Total	19	4.26787			
−1.698970	2.30103						
−0.853872	2.77815		Coefficients	Std. Error	*t* Stat	*P*-value	
−1.301030	2.47712	Intercept	−4.96055	0.18957	−26.16	8.9E − 16	
−1.096910	2.60206	*X* Variable 1	1.4805	0.0747	19.82	1.1E − 13	
−0.508638	3.00000						
−1.397940	2.47712						

The first two columns on this table are simply base 10 logs of the perfect foresight half spreads and transactions sizes, both from Table 4.1. The remainder of this table presents regression results, calculated with Excel 2007, for a test of logs of perfect half spreads as the dependent variable on logs of transaction sizes. The coefficient, *X* Variable 1 is the slope of this regression.

4.5 ALGORITHMIC TRADING

Typical transaction sizes in many markets such as the NYSE are very small relative to the order sizes placed by institutional investors. Average transaction sizes have dropped from 1477 shares in 1998 to approximately 400 shares in 2008 to below 300 in 2011. This means that the much larger institutional trades can have a significant and unwanted impact on execution prices. This unwanted price impact, which works against the trader, is referred to as *slippage*. Slippage occurs when the market impact of a trader's buy orders forces security prices to rise and sell orders force prices down. *Algorithmic trading* refers to automated trading with the use of live market data and rule-driven computer programs for automatically submitting and allocating trade orders among markets and brokers as well as over time so as to minimize the price impact of large trades. Algorithmic trading executes orders without direct human intervention, using computers to directly interface with trading venues. Algorithmic trading models seek to determine strategies for trade amounts, prices, timing, and venues of orders, in many cases, to avoid slippage. When brokers rely on algorithms to break up client orders (in this context, referred to as parent orders) into smaller slices (child orders) so as to avoid slippage, their algorithm-based trading is called *agency algo trading*. Algorithmic trading is a response to market imperfections (e.g., asymmetric information availability), and may contribute to market imperfections as well. Agency-only execution firms (e.g., CF Global Trading and WallachBeth Capital) and other brokers employ algos to implement the best executions for their clients. Many institutional traders employ algorithmic trading (also called *automated trading*, *black box trading*, and *robotrading*) or agency execution services to break up large parent orders into smaller child orders to reduce execution risk, preserve anonymity, and to minimize trade slippage. Rooted in *program trading* (defined by the NYSE as computer-initiated trades involving 15 or more stocks with value totaling more than $1,000,000), which dates back to the 1980s, algorithmic trading now accounts for between 30% and 80% of trade volume in many markets.

While algorithmic trading frequently refers to the trade execution models referred to above, the term is also used in a more general sense to include *alpha models*, which are used to make trade decisions to generate profits or control risk. More generally, algorithmic trading can be defined as trading based on the use of computer programs and sophisticated trading analytics to execute orders according to predefined strategies. Thus, algorithmic trading can be generalized to include program trading. Regardless, algorithmic trading is usually highly dependent on the most sophisticated technology and analytics.

Algorithmic trading does have risks, such as leaks that might arise from competitor efforts to reverse engineer them. Designers of algorithmic trading models seek to ensure that trading schedules and behavior cannot be predicted. However, many algorithms lack the capacity to handle or respond to exceptional or rare events. In addition, any malfunction, including a simple lapse in communication lines, can cause the system to fail. Thus, human supervision of algorithmic trading and appropriate use of filters are crucial.

Algo Strategies

Strategies used by algorithms vary widely. Sometimes, algorithmic trading results from mathematical models that analyze every quote and trade in the relevant market, identify

liquidity opportunities, and use this information to make intelligent trading decisions. This type of strategy is sometimes called *smart order routing*. Some algorithmic trading models seek to trade at or better than the average price over a day (e.g., *VWAP*, volume weighted average price, or *TWAP*, time weighted average price—both are discussed in Chapter 10) and others seek to execute slowly so as to have minimal price impact. For example, with *simple time slicing*, orders are split up and sent to markets at regular time intervals. Some strategies are more competitive, such as *iceberging* where small parts of orders are revealed to determine their impact and *pegging* where orders are sent to execute at the best bid or offer to test whether the order moves the market. Algorithms sometimes are set to produce more volume at market opens and closes (e.g., *MOC*, market on close) when volume is high, and less during slower periods such as around lunch. They can seek to exploit any arbitrage opportunities or price spreads between correlated securities.

Newsreader algorithms employ high-speed text mining and statistical techniques to analyze news sources, blogs, tweets, and other data to obtain relevant trading information and infer its impact on security prices. Essentially, data are mined and read and content is mechanically analyzed for trading decisions. More generally, *information algorithms* search and filter information and news flow in an effort to locate trading opportunities and trigger trades. These algorithms can be structured to alert human traders or structured to execute automatically based on rules provided by the user.

In addition, some algorithms seek opportunities in options and other derivatives markets. Algorithmic trading can also account for and seek to balance trading costs such as commissions and taxes as well as regulatory issues such as those associated with short selling. Algorithms are created to account for the relative importance of the tradeoff between the order's speed of execution and its price impact. Many algo strategies are highly dependent on trading venue selection and on speed. We will discuss strategies of venue selection in Section 4.6 and algorithms used by High Frequency Traders in Section 4.8.

Illustration: Slicing the Large Transaction

Earlier in this chapter, we discussed best execution, slippage, and the perfect foresight half-spread. Suppose that a client of the agency-only XYZ brokerage firm seeks to have her purchase order of $X = 80{,}000$ shares completed with the best possible execution; that is, she wants to slice her 80,000-share order so as to realize the lowest possible total execution costs B. These execution costs B will include both order processing costs F per transaction and market impact (slippage) costs that are increasing in the size c of each of n equal-size transactions. Broker XYZ works with a market impact model that estimates the slippage cost S of each of n equal-sized executions $c = X/n$ with function $S = vc^m$, where the coefficient v and exponent m are related to the liquidity of the security to be purchased or sold.[4] In this case, the per-transaction cost B/n is estimated to have a fixed component F (order processing

[4]Recall that we used a log-log regression estimation procedure to determine the values of parameters v and m. Some firms use square root, quadratic, simple linear, or other cost functions. However, significantly more sophisticated cost functions are also used, accounting for factors such as market liquidity, different markets, time of day, and the urgency of trading, and making use of more sophisticated and dynamic optimization procedures such as Euler and Bellman equations.

costs) and an increasing variable cost component $vc^m = v(X/n)^m$ (this is the market impact component of the execution costs):

$$B/n = F + v(X/n)^m$$

Suppose that broker analysts have estimated F to be \$0.03 per order, based on knowledge of order processing costs, and v and m to be 0.00001 and 1.5. What are the optimal number of transactions and transaction sizes? To answer these questions, we will set up a total transaction cost function as follows:

$$B = nF + nv\left(\frac{X}{n}\right)^m = 0.03n + 0.00001\frac{80,000^{1.5}}{n^{0.5}} = 0.03n + 226.27n^{-0.5}$$

We find the derivative of total transaction costs B with respect to the total number of transactions n, set this derivative equal to zero and solve for n as follows:

$$\frac{dB}{dn} = \frac{d\left[nF + nv\left(\frac{X}{n}\right)^m\right]}{dn} = F - (m-1)v\frac{X^m}{n^m} = 0.03 - 0.5(0.00001)\frac{X^m}{n^m} = 0.03 - \frac{113.14}{n^{1.5}} = 0$$

$$n = \sqrt[m]{\frac{(m-1)vX^m}{F}} = \sqrt[1.5]{\frac{0.5(0.00001)\cdot 80,000^{1.5}}{0.03}} = 242.28$$

Thus, we find that the optimal number of stock transactions for this order is $n = 242$, each with a transaction size of X/n equal to 80,000/242 = 331 (subject to rounding).

The algorithm that we have just illustrated is very simple. Many potential enhancements could be developed to enrich the model. For example, the model provides no guidance on spacing trades through the day or time frame. First, we might expect that the amount of time elapsing between trades executing at time t and time $t+1$ would be inversely related to market impact costs. That is, as more time elapses between the broker's transactions, slippage will be reduced. Second, many securities might be expected to be more liquid during the earliest and latest parts of the day, and least liquid in the middle. That is, one might expect liquidity during the day to be "U"-shaped, and market impact costs to resemble an upside-down U. Trade urgency, competing market venues are among the many other potential factors were not considered here. Clearly, the analyst developing algorithmic programs has many factors that she can consider when designing the optimal cost or trade-slicing functions.

4.6 DARK POOLS

Sometimes, all or parts of orders generated by algorithmic traders are hidden from the eyes of the public as opposed to orders quoted in "lit" markets such as exchanges. Lit markets display quotes and execution prices to the public while dark pools display only prices after order executions to the consolidated tape, and then normally, only after a delay. These intentionally obscured portions of orders are generally referred to as *dark liquidity pools* because they are hidden from public view until after they execute. These dark pools

help traders with large order sizes maintain anonymity and avoid slippage while attempting to execute their large transactions. Orders are often partially revealed, in which case they are called *icebergs* (hidden-size orders are generally completely hidden), with brokers instructed not to reveal the full size of their orders. Hidden portions of orders do not contribute to the display of market depth. In many markets, the fully revealed portion of the order will normally have priority over the hidden portion. Mittal (2008) reported that dark pools represent about 7% of total equity market volume in 2006; more recently, this figure has risen to about 15%.

Dark pools also refer to alternative trading systems (ATSs) that do not post quotes and transaction prices in the traditional sense. In some cases, this information (often referred to as alerts) can be transmitted to certain traders such as those who can potentially provide for a match; in other cases, information is tightly restricted. These dark pools function in a parallel fashion with traditional markets, executing transactions within the national market bid/offer (NBBO) spread, usually at the NBBO mid-point. Citadel and KG Holdings Inc. are among the firms that offer such confidential execution services. The Investment Technology Group (ITG) and Mittal (2008) characterize five basic types of dark pools:

1. *Public Crossing Networks*: Usually created by brokerage firms to act on an agency basis for the purpose of generating trading commissions on large trades, such as ITG's POSIT and Liquidnet. They normally cross buy and sell orders, frequently at the spread mid-point.
2. *Internalization Pools*: Usually created by brokerage firms for internalizing the broker's trade flow, such as Credit Suisse's Crossfinder, and Goldman Sach's Sigma X (operated by NASDAQ as of late 2016). Internalization of order flow means that the broker provides both the buy and sell flow from its own proprietary desk and its customers, maintaining complete control over the transaction and reducing transaction costs.
3. *Ping Destinations*: These pools, operated by hedge funds and electronic market makers, accept only immediate or cancel (fill or kill) orders. Such markets usually provide for low transaction costs and can facilitate flash trading (discussed in the upcoming section on high-frequency trading).
4. *Exchange-Based Pools*: ATSs created by exchanges, such as ISE Midpoint Match, NASDAQ Cross, and NYSE Matchpoint. In many cases, suppliers of liquidity will receive rebates for providing liquidity, while buyers of liquidity will pay higher commissions for the liquidity that they receive.
5. *Consortium-Based Pools*: Operated by multiple broker partners and function as a hybrid of public crossing networks and internalization pools, such as LEVEL and BIDS. In a sense, a consortium-based pool acts as a sort of second-step internalization pool. If a broker cannot internalize an order, it seeks to execute the order within its consortium of partners before sending it out to the public.

Trader anonymity and reduced slippage are clear advantages to trading in a dark pool. Most trades are executed within the bounds of the NBBO. While anonymity is essential to the dark pool trading venues, traders have learned to game the pools. For example, in Chapter 13, we will discuss *fishing*, the practice involving gamers sending small orders to dark pools to determine whether they can be executed, in which case they often infer that larger numbers of shares await in the pool. Gamers can use this information to set their own trading strategies and game the market. Fishing activities obviously impact market

prices, as does the removal of orders from display markets to dark pools. If dark pools drain significant volume from public display markets, the price discovery process might become significantly impaired, and market efficiency (the ability for markets to fully reflect available information) would be adversely affected.

Menkveld, Yueshen, and Zhu (2017) introduce a theory of and offer empirical support for pecking orders for trading venues. They argue that investors rank trading venues based on low-cost-low-immediacy versus high-cost-high-immediacy. When trade urgency is low, low-cost mid-point dark pools rank first, followed by other dark pools and lit (high-transparency) markets such as exchanges at the bottom. Lit markets are the most costly in that they provide the least protection against information leakage and slippage. However, when trade urgency is high, investor rankings are reversed, with lit markets at the top despite their higher information and slippage costs and mid-point dark pools at the bottom due to execution uncertainty. Thus, for example, when market volatility increases, or when there are earnings announcement surprises, trade urgency increases and lit markets become the preferred trading venue due to their certainty of execution.

4.7 STEALTH AND SUNSHINE TRADING

As we discussed earlier in the context of algorithmic trading, investors with private information should trade gradually to disguise their intentions and to realize trading profits before their trades fully reveal valuable information. They fragment their orders, routing them to different markets and at different times to disguise their intentions. This is the purpose of stealth trading.

To motivate discussing the need for stealth trading to avoid slippage, consider the extent to which small, uninformed investors cause stock prices to change versus the effects of large informed traders pushing prices towards their fundamental values. Barclay and Warner (1993) argue that if informed investors' trades are the main cause of stock price changes, large trades by informed traders would reveal their information because traders can easily discern those who execute large transactions. However, small trades executed repeatedly will incur excessive transaction costs. Thus, informed traders will tend to fragment orders into medium-sized trades sufficiently small to disguise themselves but large enough to maintain fairly small transaction costs. This should lead medium-sized transactions to have the most significant impact on stock prices. Thus, medium-sized orders (500 to 9999 share orders) will be most likely to move stock prices if stealth traders are most active. Several empirical studies support this argument, including that of Barclay and Warner who find that 99.43% of the cumulative price changes on the NYSE result from medium-sized trades, which comprise 38.12% of total NYSE trades. Chakravarty (2001) confirms that the bulk of price changes are due to medium-sized trades, finds that the bulk of these price-moving trades are executed by institutions, and determines that the bulk of stealth trades are executed by institutions. Chakravarty, Kalev, and Pham (2005) found that institutional investors appear to be more informed than individual investors, enabling them to function more appropriately as stealth traders.

In contrast, Campbell, Ramadorai, and Vuolteenaho (2009) found that institutions tend to execute larger and smaller transactions rather than medium-sized transactions. This

differs from earlier studies cited above that suggest that institutions execute medium-sized transactions. However, the Campbell et al. (2009) study was more recent and the exchanges have consistently been reporting smaller average transaction sizes. Perhaps significantly reduced transaction costs have made execution of smaller transactions less costly, leading institutions to execute an increased number of smaller transactions. Furthermore, Campbell et al. also found that institutions prefer medium and small size trades on high volume days and larger size trades on high volatility days. Furthermore, larger transactions move prices more on high volatility days. This suggests that stealth trading increases when liquidity is high, but when volatility is high, traders act more quickly with larger size trades, engaging less in stealth trading, causing their trades to move prices more.

Chakravarty et al. (2005) found that stealth trading occurs differently in bear markets from how it occurs in bull markets. While they did find that informed traders execute smaller trade sizes when selling than when buying, informed investors fragmented their trades more when buying in rising markets and when selling in falling markets. Furthermore, some stealth trading algorithms seek opportunities in options and other derivatives markets. In fact, Chakravarty, Gulen, and Mayhew (2004) demonstrate that informed traders trade across both stock and options markets and that a significant level of price discovery occurs in options markets.

Sunshine trading is the opposite of stealth trading, as sunshine traders announce their intentions. The sunshine trader announces its intentions when doing so might be expected to increase the competition to act as counterparties on the same transactions, but when additional competition on the same side of the transaction won't hurt the institution. Sunshine trading is most effective when the market for the security is elastic, that is, when a larger number of competitors on the same side of the transaction might be expected to have little long-term impact on the price of the security. Even if the announcement produces a short-term impact, in an elastic market, the price will bounce back. If information motivating a trade has already been leaked, the institution might publicize its intentions in an effort to obtain a larger number of providers of liquidity. That is, the institution might prefer to execute its transactions in a more liquid market. If forthcoming trades are not motivated by information, preannouncing the motivations of the trade may result in better trade prices than stealth trading. In fact, the signaling effect of sunshine trading might actually "scare off" potential competitors on the same side of the transaction, leading to reduced trading by competing informed traders.

Consider, for example, the 2011 scenario when the U.S. government announced its intent to sell bank shares accumulated in its efforts to mitigate the effects of the 2008 banking crisis. The market was well aware of the U.S. government's holdings and its intent to eventually sell the shares. Hence when these shares ultimately were sold, the government's announcement of the anticipated sales prevented suspicions leading to unwanted negative information effects.

4.8 HIGH-FREQUENCY TRADING

High-frequency trading refers to the practice of rapid executions of multiple transactions for securities followed by extremely short holding periods, perhaps as short as

fractions of seconds. High-frequency traders typically use their technology and trading algorithms to exploit market inefficiencies in market making, price effects of news, and market movements. High-frequency traders focus on small price changes and discrepancies, often "ticks." Obviously, high-frequency trading requires sophisticated technology, communications, and computing resources costing millions or even tens of millions of dollars. Several studies have indicated that high-frequency trading accounts for more than half of all U.S. equities trading volume and over 45% in Europe, 40% in Japan, 27% in Australia, and 12% in the rest of Asia as of the end of 2012 (Australian Securities and Investments Commission, 2015; Popper, 2012). This means that high-frequency trading has the potential to add liquidity to the marketplace. High-frequency trading obviously depends heavily on execution speed. In fact, the prevalence of HFT implies that at least half of all trades are not motivated by liquidity, the need for portfolio diversification, fundamental analysis, or technical analysis. These trades are motivated by latency. Many markets measure *latency*, the delay between an order signal and its execution in milliseconds. For example, the NYSE SDBK system has reduced latency to approximately 5 milliseconds (0.005 seconds) for most of its high-frequency transactions. KCG LLC is one of the major firms focused in the high-frequency execution services industry, and hedge funds such as Citadel Investment Group and Renaissance Technologies LLC are also major participants as HFT traders. Wah (2016) estimates the 2014 industry-wide total profits for HFT trading to be $3.03 billion.

In recent years, major commercial and investment banks such as Goldman Sachs and have lost ground in the HFT arena to smaller proprietary trading firms and hedge funds that specialize in the practice, including Virtu Financial, KCG, Hudson River Trading, and Jump Trading. Nonetheless, a high-frequency program requires significant data, technological, and personnel resources. In addition to the ability to obtain, transmit, and analyze the most recent and complete quote and execution data, a high-frequency trader (HFT) needs to be able to develop and implement trading strategies. These strategies will require historical tick data (trade-by-trade price data) along with quote data for testing in a data repository and data-cleansing capabilities. Second, appropriate personnel and tools need to be available to develop and test appropriate trading models under real-world conditions (see Sections 12.3 and 12.5). Such testing capabilities will include back-testing models. Third, the high-frequency trading system will require highly reliable trade-execution services that will execute on signals as prescribed by model developers.

Some HFTs, specifically *liquidity rebate traders*, rely heavily on payment for order flow (introduced in Section 2.4). Exchanges and many electronic communications networks (ECNs) offer rebates of about ¼ penny per share to certain broker dealers that execute orders. These executed orders supply liquidity in the market, and the payments or rebates for this order flow might be regarded as payment for supplying liquidity. Brokers and traders using this liquidity might be expected to pay higher fees to the exchange or ECN. These rebate programs have led to trading strategies designed to capture these rebates. A dealer purchasing shares at 50.00 and quickly reselling them for 50.00 may still make a ½ penny profit per share from the liquidity rebate. Buy-side traders pay this fee through higher exchange/ECN fees, but these fees might be regarded as the cost of using liquidity in the market. Some market centers maintain an explicit pricing scheme called *maker-taker* pricing: brokers who make liquidity through limit orders receive a payment; takers of

liquidity through market orders pay a small fee. This pricing system is designed to encourage market participants to supply orders that enhance the liquidity of the market. We will discuss controversies associated with payment for order flow in Section 13.5.

Interestingly, high-frequency trading not only relies heavily on algorithms to provide rules and strategies for trades, but many algorithms are designed specifically to anticipate and exploit the algorithms used by other institutions. Hence, some HFT algorithms are designed to detect ("sniff out") and exploit competitors' algorithms. For example, a small sell order for 100 shares on a particular exchange placed by a given broker might imply a much larger latent sell order awaiting execution. Sniffing algorithms are designed to detect such latent orders.

High-frequency trading has gained significant importance, particularly in equity and exchange markets. Major brokerage houses, trading firms, and hedge funds often invest tens of millions of dollars into their high-frequency trading technologies and capabilities. For example, many firms engaging in high-frequency trading have co-location facilities, where they rent space from market centers, such as from the major exchange ECNs and other ATSs. The purpose of co-location where their servers are located next to the market center's data servers is to reduce latency. This co-location (also called proximity hosting) provides significant millisecond and even microsecond advantages, and enables institutions to better serve their clients and obtain and react to competitors' trading behavior more quickly.[5] However, the fairness of this high-cost practice has been questioned and remains under SEC scrutiny.

Latency Arbitrage

Latency is concerned with the speed of quotation display or order execution after an order has been placed. Different technologies and resources (such as co-location, locating trader computing and order routing facilities in the buildings where markets are located) provide opportunities for some traders to obtain quote and order information faster than others. For example, HFT SIPs simply process faster than the SIPs maintained by the exchanges on which customer orders are based. This HFT speedy access to information enables them to exploit their information advantage over their competitors through *latency arbitrage*. Arbitrage is a series of simultaneous or nearly simultaneous transactions to exploit price differences in securities or portfolios producing identical or nearly identical cash flows across multiple markets. Latency arbitrage typically involves an HFT trader "picking off" orders in one market before other traders have a chance to respond to orders in another market, sometimes offsetting the transaction in another market.

Exchanges disseminate regular updates of their best bids and offers; NBBOs on stocks are compiled from these and routed to clients. However, many dark pools, or anonymous

[5]Consider the following: In 1 microsecond, light (or, an electronic signal) can travel about 300 meters. Thus, it takes at least 200 microseconds for an electronic signal to travel 60 kilometers from the NYSE in Manhattan to a hedge fund in Greenwich, Connecticut. This calculation ignores the additional latency induced by routers, switches, and so on. Obviously, when time is of the essence, location matters. In fact, some markets provide institutions that co-locate in their facilities with cables of identical length to connect data feeds, regardless of how many meters the client computers are from the exchange computers.

electronic markets such as Liquidnet that don't publish bids and offers, don't contribute to the NBBO. Delays between the appearance of bids and offers on direct feeds and the processing of the NBBO provide arbitrage opportunities as do the abilities of institutions to synthesize NBBOs before they are even available to most traders. Consider the following illustration of a HFT whose automated trading facility enables it to exploit its information advantage:

1. The initial market for Stock X is 50.00 bid/offered at 50.01.
2. The HFT's superior technology and location enables it to pick up a new institutional bid for 50.01 for Stock X that is deemed likely to be followed by more and higher bids. This bid arrives in the HFT's computer 10 microseconds before most competing traders' computers.
3. The HFT's computer buys all available shares offered at 50.01 before competing traders act.
4. The institutional buyer is unable to purchase at 50.01.
5. The NBBO for the stock changes to 50.01 bid/offered at 50.02.
6. The HFT sells shares to the institutional buyer for 50.02, locking a 0.01 profit per share.

While the profit seems small, HFTs seek profits on a high volume of transactions. In addition, their profits may well be enhanced by market rebates.

Consider a second example described in the *Wall Street Journal*. In July 2009, Intel, the computer chip giant, reported robust earnings and some traders considered the semiconductor company Broadcom to be a buying opportunity. Many of these traders, to disguise their interest, split their orders into small batches. Broadcom began trading that day at $26.20. HFTs with faster access to quotes were able to see that there was a strong interest in Broadcom. The HFTs began buying Broadcom shares and then reselling them to the slower investors at higher prices. The price of Broadcom quickly rose to $26.39 as HFTs began offering to sell hundreds of thousands of shares. The *Wall Street Journal* reported that slower-moving traders paid $1.4 million for about 56,000 shares, or $7800 more than if they had been able to move as quickly as the HFTs.

Other HFT Strategies

Many HFT strategies, including the latency arbitrage strategy illustrated above, are quite simple. For example, many HFT desks are largely replacing the more traditional dealers as market makers by relying on simple *passive market-making*, placing quotations in large quantities. Here, the HFT strategy seeks to profit through a simple high volume based on having its own bids and offers taken. In many instances, the firm may be using its quotations to seek liquidity rebates (payment for order flow, in this instance, known more specifically as liquidity rebating) by providing liquidity in markets' maker-taker models of pricing. As discussed earlier, these markets pay HFT firms for limit orders that they can cross against counterparty market orders.

Other HFT strategies involve seeking profits through cross-asset or cross-market arbitrage strategies or by accessing information or executing trades more quickly than their competition. Co-location, the most rapid data feeds and trade execution abilities, are crucial here.

Many HFT desks seek to anticipate order flow (including sniffing out competing algorithms, a practice discussed in Chapter 13), and to use superior technology to trade ahead of this flow. For example, by studying the markets for particular securities, their trading histories, the ways that orders are placed in the various exchanges and ATS's, HFT's can often anticipate which small orders are merely parts of larger orders. The HFTs then seek to trade ahead of the anticipated slices of larger orders, in many instances, selling at a profit to the traders whose trades they sniffed out. Sometimes, these order anticipators are referred to as "parasitic traders" in that they use traders' and markets' costly information acquisition without contributing to price discovery or enhancing liquidity themselves.

There are a variety of other HFT strategies. Many simply seek arbitrage opportunities trading offsetting positions in EFTs and underlying securities. Even though most HFTs seek to hedge or neutralize their positions (be flat by the end of the trading day), some do seek profits through speculation or by making other directional trades.

Trading firms can engage in high-frequency trading by obtaining direct market access (DMA) to exchanges and other appropriate trading venues and by obtaining hardware and software necessary for strategy implementation. Software and even strategies can be obtained through brokerage firms, purchased through consulting firms and other developers, or created in-house. For example, *InfoReach* is a provider of broker-neutral software and systems for electronic, algorithmic, and high-frequency trade analysis, management, and execution through its *HiFREQ* algorithmic engine. It is designed to enable traders to create and deploy proprietary, complex trading strategies as well as access algorithms from brokers and other third-party providers. InfoReach also supplies its internal *FIX Engine* as a routing system.[6]

HFT Impact on Market Quality

Market quality might be gauged by the ability of investors to quickly access and trade based on new information in liquid markets. Clearly, equity markets are more liquid and transactions costs are lower than they were 15 years ago before the advent of HFT. While this does not necessarily demonstrate a causal effect between HFT and improved market quality, the volume of trades associated with high frequency trading certainly has the potential to increase liquidity and reduce trading costs. Of course, by definition, HFT is associated with increased trading speed. In fact, substantial empirical evidence does suggest that HFT, in general, increases market liquidity (e.g., Hasbrouck & Saar, 2013), reduces trading costs (Harris, 2013), enhances price discovery (Conrad, Wahal, & Xiang, 2015), and reduces trading spreads (Brogaard, Hagstromer, Norden, & Riordan, 2015). Brogaard, Hendershott, and Riordan (2014) find that HFT firms enhance price discovery by trading in the direction of permanent price changes and in the opposite direction of transitory pricing errors, which might have the effect of decreasing market volatility. Based on his review of approximately 30 empirical papers, Jones (2013) summarizes that HFT in competing markets does "improve market quality, reduce[s] trading costs, and make[s] stock prices more efficient."

However, this empirical evidence has not been entirely supported by other studies. For example, Kirilenko, Kyle, Samadi, and Tuzun (2017) found that during the extreme

[6]For details, see http://www.inforeachinc.com/aboutus.php.

volatility period known as the "Flash Crash" in May, 2010 (see Chapter 13: Trading Gone Awry, later in this book), HFT's did draw down liquidity as the crash worsened. Loss of liquidity when the market needs it most is known as market fragility; its opposite is market resilience or robustness. While such liquidity draw downs probably did not cause this particular crash, liquidity draw downs can exacerbate volatility in an already fragile market.

There is no question that spreads have decreased during the period of HFT existence, to roughly .003% in equity markets today relative to .9% in the mid-1990s. But, how much of this spread reduction is due to HFT, and how much of this reduction is real rather than illusory? Most quotes can be withdrawn nearly instantly by HFT traders, which renders many quotes rather meaningless if they cannot be taken by other investors. Over 95% of quotes in stock markets are ultimately cancelled. For example, there is evidence that some market liquidity and market spreads are illusory (Eggington, Van Ness, & Van Ness, 2016; Coffee, 2014), in large part due to flash trading and quote stuffing, the potentially unethical practice of placing and withdrawing large numbers of quotes (see Chapter 13: Trading Gone Awry, for details on these and other problematic practices often associated with HFT shops).

Most HFT activity is based not on costly and time-consuming fundamental research but on simply observing and anticipating the order flow. That is, HFT traders do not normally conduct the costly and time-consuming firm-based research needed to ensure efficient market pricing. Instead, some observers (e.g., Brogaard, Hendershott, & Riordan, 2014) argue that many HFT traders free-ride on the costly efforts of analysts and investors by observing and anticipating the order flow of informed investors, and then use their superior technology to advance ahead of the informed trader order flow. Eventually, the inability of slower informed traders to compete might be expected to reduce their incentives to trade, thereby reducing their willingness to conduct the valuable research that is required for market efficiency and optimal capital allocation.

HFT is a drain on trader resources. The largely socially wasteful "technological arms race" (Budish, Cramton, & Shim, 2015; Section 1.8 earlier in this book) associated with HFT imposes a substantial cost on the HFT trader that seeks the fastest response to new information. In addition, it imposes an externality on the rest of the industry as each HFT competitor makes similarly large investments to win the race (winning, not placing second or third is key in an efficient market), resulting in a suboptimally large industry investment in technologies ultimately rendered obsolete. Winning the race is not the only issue for traders; they need extraordinary speed to avoid having their stale quotes (quotes not based on the newest information or transactions) picked off (taken by faster traders before they can be canceled). Evidencing this assertion of socially suboptimal investments in faster technology, the Tabb Group estimated that trading firms spent approximately $1.5 billion in 2013 alone to increase trading speeds (Patterson, 2014), also supporting the financialization argument in Section 1.8 of this book. While HFT has enjoyed a profitable run, numerous sources (e.g., McNamara, 2016) observe that a number of HFT shops are closing as profitability drops and risks remain high.

Reaction to HFT: The IEX

The Investor's Exchange (IEX) was recognized by the SEC to operate as an exchange in June 2016, after functioning as an ATS as of October 2013. The IEX claims that its rules

and technology impede latency arbitrage and "electronic front running," thereby offering "simple, fair and transparent markets" for its participants.[7] The exchange was designed to impede the HFT's ability to exploit slower traders. The IEX forgoes the "fast lane, slow lane" model of order executions, slows down transaction executions and reduces broker conflicts of interest by

- prohibiting colocation with the IEX
- refusing to make payment for order flow, instead charging a flat fee of $.0009 per share traded
- installing a "speed bump" with approximately 38 miles of coiled cable in front of its order processing system to slow down order entry by a minimum of 350 μs and
- accepting only a small number (5) of specific order types: market orders, limit orders, primary peg orders (described below), midpoint peg orders (orders that execute at the midpoint of the NBBO, subject to any limit) and its own discretionary peg orders (described below).

The speed bump is intended to protect IEX traders from having their orders picked off by HFT traders when the NBBO is transitioning. Without the speed bump, the fastest traders can react to a change in the NBBO, which might be the change in a quote in another market, and then pick off a slower trader's stale quote placed in the IEX before the slower trader has an opportunity to respond to the new NBBO. The delay also allows a large trader executing on the IEX sufficient time for the remainder of its orders to arrive at other exchanges before other market participants (e.g., HFT) discover the IEX execution and react.

After contesting the registration of the IEX with the SEC as an exchange on the grounds that its speed bump conflicts with rules that require price quotes to be immediately accessible, the NYSE has initiated a somewhat similar speed bump on its NYSE American Exchange for smaller issues. The much smaller Chicago Exchange has also installed a speed bump.

Peg orders or quotes are pegged at the NBBO and rise and fall with the NBBO. The IEX primary peg order is essentially a nondisplayed limit order pegged to the NBBO that executes at or better than the client's unrevealed limit price. That is, unlike typical limit orders, this limit (bid or offer) does not display as part of the primary pegged order. In effect, the order displays at the NBBO rather than at its limit. Because the primary peg order limit is not displayed, its entry to the order book reveals less limit order information leading to predatory behavior by other traders before it executes. When the IEX "Crumbling Quote Indicator" indicates that the security price is stable, primary pegged orders can execute at the NBB or NBO if the limit condition is fulfilled. However, if the Indicator indicates that the price is unstable, the order will execute only at a price one tick less aggressive than the NBB or NBO, assuming that the limit condition is fulfilled.

Complicating matters, the NBBO is not a constant, nor does it reflect the best quotes for all traders with different order sizes in different markets. First, a transitioning NBBO might render a slower trader's quote stale (a stale quote does not reflect the latest information) and vulnerable to being picked off by HFT traders. The discretionary peg order acts as a primary peg order or, when an IEX algorithm (the IEX Signal, also known as the "Crumbling Quote Indicator") identifies a scenario characterized by a transitioning NBBO or crumbling quotes

[7]See https://iextrading.com.

(identical quotes on multiple exchanges at the NBBO, but, one at a time, rising or falling to a less aggressive range – falling bid or rising offer, until the NBBO itself becomes less aggressive), the algorithm triggers an execution algorithm set by the IEX.[8] This execution algorithm adjusts as the NBBO and certain other quotes (e.g., quotes that might be competitive with the NBBO given their order sizes or markets) varies. This discretionary peg order is intended to foil predatory behavior by HFT traders. Other exchanges have developed or are developing variants of this order type. For example, the NYSE American Exchange and the TSX Alpha Exchange have imposed delays only on liquidity takers. In some cases, the exchange delays are for fixed periods (e.g., IEX, NYSE American) and in others, they are randomized (e.g., TSX Alpha, Eurex Exchange, Germany).

The overall impact of IEX innovations is not yet clear. As we have discussed, it is clear from the more general empirical literature that HFT has reduced spreads and reduced transactions costs in equities markets. However, IEX supporters argue that most of these gains have accrued to exchanges and to institutional traders. Empirical findings by Chow, Lahtinen, and Lawrey (2021) suggest that the IEX speed bump has further reduced realized half-spreads and trading costs, benefiting individual investors. On the other hand, Aoyagi (2019) argues that poorly designed speed bump installations can backfire, widen spreads, and lead to increased HFT domination. Shkilko and Sokoloc (2020) find that temporary weather-related disruptions in microwave communication networks (see Section 1.8 earlier for a related discussion) that slow HFT reduce both picking-off activity (stale quote-related exploitation and adverse selection) and trading costs, suggesting that "natural speed bumps" might serve as a short-lived protection for non-HFT investors.

4.9 FLASH TRADING AND SPONSORED ACCESS

The *Quote Rule*, formally referred to as SEC Rule 602 of Regulation NMS, requires that all market centers publicly disseminate their best bids and offers through the securities information processors (such as the Consolidated Tape, Consolidated Quotations System, and the Intermarket Trading System discussed in Section 2.6). However, an important exception to this Rule involves quotes that are immediately executed or canceled, defined as occurring within 500 milliseconds of the original quote. These quotes need not be publicly displayed, but may be displayed only to select clients. Such displays are referred to as *flash orders*. *Flash trading* refers to trading motivated by private or select market participant receipt of quotes (flash orders) in advance of the public quote stream (also called pre-routing displays). Most exchanges have facilitated flash trading at some point, although several have stopped the practice. Flash orders, in effect, briefly (up to 500 milliseconds, but often for as few as 30 milliseconds) display an order that is better than the national best bid or offer (NBBO) in order to solicit NBBO-priced responses. Traders requiring the fastest execution may respond, but the flash trader obtains market information from the

[8]Suppose, for example, the bids for XYZ stock on NYSE, Bats and the IEX are all 10.00, which is the National Best Bid (NBB). The NBBO reflects crumbling quotes if the bid on IEX drops to 9.98, followed by Bats to 9.98 and then by NYSE to 9.98. These bids are less aggressive than before. In response to these crumbling quotes, the IEX triggers its execution algorithm.

flash order regardless. In addition, the flash trader might be enhancing the liquidity of the market, perhaps enabling herself to secure a market trading rebate (payment for order flow). In fact, one might argue that flash orders merely provide an exchange of an opportunity to reduce trading costs for investors by executing on that particular exchange rather than distributing the order to the wider market. Furthermore, most broker clients can opt out of having their orders flashed by informing their brokers of their wishes.

On the other hand, the NYSE and others have argued that flash orders essentially provide private quotes for select participants competing with the public quote stream, impairing the importance of the NBBO and weakening the market by detracting the pursuit of best execution or best price. While the NYSE did not explicitly permit flash trading prior to 2010, some observers noted that NYSE specialists had the opportunity to act on quote information in advance of the public, although this seems like less of a problem since they were replaced by designated market makers. The SEC and the U.S. Congress have considered banning flash trading. For instance, in its 2009 proposal to ban flash trading, the SEC stated that it was concerned flash orders might create a two-tiered market in which the public does not have best price information access through the consolidated quotation data streams.

But, flash orders and flash trading might provide opportunities for traders who do not want to make their interest public to participate in the market, thereby enhancing liquidity of that market. In addition to providing enhanced liquidity, the flash trader may simply be satisfying other trader needs for immediacy. While the trader who places a limit order might be regarded as a patient provider of liquidity, the trader who places a flash order might be regarded as an impatient provider of liquidity. Either may serve to improve markets for long-term investors, particularly through the potentially reduced costs of liquidity. Nevertheless, flash orders do provide the flash trader special access to initial trader's order information as we discuss in the illustration below, potentially impairing the market as we discuss in Chapter 13.

Illustration: A Flash Trade

At 10:00.00.00, a pension fund submits a buy order for 10,000 shares of ABC stock. The NBBO at this time is 10.00−10.05, both quotes on Nasdaq, the offer for 1000 shares. The pension fund's broker routes the order to Nasdaq, which has the best price, and at 10:00.00.10, the pension fund purchases 1000 shares for 10.05. No more shares are available at this 10.05 offer as Nasdaq's best offer increases to 10.10. The NBBO changes to 10.02−10.07, with the best offer now being for 9000 shares on Direct Edge at 10.07. Before routing the order to Direct Edge where the best offer is, the pension fund's broker flashes for 100 milliseconds the order to select Nasdaq HFT traders at 10.06, slightly better than the best offer on Direct Edge. Nasdaq HFT Flash order recipients have an option to react to this flash order by selling 9000 shares at 10.06.

Scenario 1: Flash order recipients sell 9000 shares of stock to the pension fund by taking its bid for 10.06. In this scenario, the pension fund saves \$90 ($9{,}000 \times (\$10.07 - \$10.06) = \$90$) relative to taking the 10.07 Direct Edge offer for its 9000 shares.

Scenario 2: Suppose, instead, Nasdaq HFT traders exposed to this flash order react by withdrawing all their offers (Note that Arnuk and Saluzzi (2012) report that approximately 95% of all quotes are canceled) and by nearly instantly buying all 9000 available shares on

Direct Edge for 10.07. At 10:00:00.60, HFT traders are offering their shares at 10.09 on Direct Edge, with the NBBO now at 10.05 – 10.09, with the 10.09 offer on Direct Edge for 9000 shares. The pension fund's order is routed to Direct Edge, which has the best price, where the fund buys all 9000 shares from HFT traders for 10.09. We see in this scenario that HFT traders made a profit of \$180 ($9000 \times (\$10.09 - \$10.07) = \$180$) in less than one second by acting on the advance flash order information. Correspondingly, the pension fund paid an additional \$180 due to its order being flashed.

Clearly, the first scenario above would have been preferred by the pension fund. Actually, this scenario has the possibility of working to the pension fund's further advantage in a maker/taker payment system; the fund stands to receive a liquidity rebate (unless its broker takes it) should the order be taken at the flash. However, the second scenario, with the share demand being flashed to HFT traders is an example of why flash trading has caused so much controversy, as we will discuss further in Chapter 13. Of course, almost nobody likes a market in which one tier of traders seems to have a consistent informational advantage over everyone else.

Sponsored Access

Another concern to many high-frequency trade critics is the practice of registered broker-dealers providing *sponsored access*, where brokers loan their *Market Participation Identification Numbers* (MPIDs) to HFTs, enabling them to gain faster, direct access to markets by bypassing the broker's trading systems. Normally, brokers and traders must be registered members or licensed to have direct access to a trading venue such as an exchange. Sponsored access enables nonmembers to access trading venues. *Naked access*, where orders do not pass through the sponsor's pretrade risk controls, is of particular concern, since such algorithmic orders could lead to runaway executions.[9] Naked access might speed the transaction execution from as much as 750 microseconds when the trade is routed through risk controls to as little as 250 microseconds with unfiltered access. Tracking down destabilizing naked access trades run amok can be tricky because exchanges cannot know where they originated, and the sponsoring brokerage firm must be involved in the identification. While as much as two thirds of equities volume may involve high-frequency orders, perhaps as much as half of these trades might involve sponsored access. In 2011, the SEC implemented a ban on naked access, expressing concerns about "the potential that financial, regulatory, and other risks associated with the placement of orders are not being appropriately managed. In particular, there is an increased likelihood that customers will enter erroneous orders as a result of computer malfunction or human error, fail to comply with various regulatory requirements, or breach a credit or capital limit."[10] We will discuss problems originating with "unfiltered" trading in Chapter 13.

[9]In November 2010, largely in reaction to the May 6, 2010 "flash crash" (the "flash crash" had nothing to do with flash trading), the SEC voted for a ban on naked access in stock markets, requiring brokers "to put in place risk management controls and supervisory procedures to help prevent erroneous orders, ensure compliance with regulatory requirements, and enforce pre-set credit or capital thresholds." In 2011, this ban was implemented as per Market Access Rule 15c3-5.

[10]From a January 13, 2010 SEC press release, available on the SEC website, http://www.sec.gov/news/press/2010/2010-7.htm.

Some market observers argue that high-frequency trading enhances market liquidity, bringing into the market pools of securities when other traders need trade counterparties, and providing over half of the total volume in some markets. However, many are also concerned about risks of this high-volume, "trigger-happy" trading, perhaps fed by algorithmic, technological, or trading errors that could trigger massive one-sided order flows leading to failures, contagion, and market meltdown.

Additional Reading

Harris (2003), Chapters 13 to 18, is a very useful guide for understanding market mechanisms and trading institutions, although it has become somewhat dated. Kim (2007) offers excellent explanations of electronic and algorithmic trading technology in Chapters 1 to 3 and 6 to 7. Schwartz (1988) provides a dated but very readable discussion focused on equity markets in Chapters 1 to 15. Lewis (2014) is a very lively and popular book concerning the conflicts between financial industry participants, particularly those engaged in high frequency trading, and the investing public. The book introduces dark pools, HFT and algo trading in a nontechnical and very readable fashion.

References

Aoyagi, J. (2019). *Strategic speed choice by high-frequency traders under speed bumps*. ISER Discussion Paper 1050. Osaka, Japan: Institute of Social and Economic Research, Osaka University.

Arnuk, S., & Saluzzi, J. (2012). *Broken markets: How high frequency trading and predatory practices on Wall Street are destroying investor confidence and your portfolio*. Upper Saddle River, NJ: FT Press.

Australian Securities and Investments Commission. (October 2015). *Review of high-frequency trading and dark liquidity*. Report 452.

Barclay, M. J., & Warner, J. B. (1993). Stealth and volatility: which trades move prices? *Journal of Financial Economics*, *34*, 281–306.

Boston Consulting Group. (2021). *Global asset management 2021: The $100 trillion machine*. Available at <https://web-assets.bcg.com/e2/d1/e3de3ed0408ba1b0f269f33d4d97/bcg-global-asset-management-2021-jul-2021.pdf> Accessed 20.09.21.

Brogaard, J., Hagströmer, B., Norden, L., & Riordan, R. (2015). Trading fast and slow: Colocation and liquidity. *Review of Financial Studies*, *28*, 3407–3443.

Brogaard, J., Hendershott, T., & Riordan, R. (2014). High frequency trading and price discovery. *Review of Financial Studies*, *27*, 2267–2306.

Budish, E., Cramton, P., & Shim, J. (2015). The high frequency trading arms race: Frequent batch auctions as a market design response. *The Quarterly Journal of Economics*, *130*(4), 1547–1621.

Campbell, J. Y., Ramadorai, T., & Vuolteenaho, T. O. (2009). Caught on tape: Predicting institutional ownership with order flow data. *Journal of Financial Economics*, *92*, 66–91.

Chakravarty, S. (2001). Stealth-trading: Which traders' trades move stock prices? *Journal of Financial Economics*, *61*, 289–307.

Chakravarty, S., Gulen, H., & Mayhew, S. (2004). Informed trading in stock and options markets. *Journal of Finance*, *59*, 1235–1257.

Chakravarty, S., Kalev, P., & Pham, L. (2005). *Stealth trading in volatile markets*. Working paper. Melbourne, Victoria: Department of Accounting and Finance, Monash University.

Chan, L. K. C., & Lakonishok, J. (1993). Institutional trades and intraday stock price behavior. *Journal of Financial Economics*, *33*, 173–199.

Chow, A., Lahtinen, K. D., & Lawrey, C. (2021). The Investors Exchange's (IEX) impact on investors. *Journal of Asset Management*, *22*, 51–61.

Christie, W. G., & Huang, R. D. (1994). Market structures and liquidity: A transactions data study of exchange listings. *Journal of Financial Intermediation*, *3*, 300–326.

Coffee, J. C. (2014). *High frequency trading reform: The short term and the longer term*. The CLS Blue Sky Blog, July 21, 2014 <http://clsbluesky.law.columbia.edu/2014/07/21/high-frequency-trading-reform-the-short-term-and-the-longer-term/> Accessed 02.07.17.
Conrad, J., Wahal, S., & Xiang, J. (2015). High-frequency quoting, trading, and the efficiency of prices. *Journal of Financial Economics*, *116*, 271–291.
Dennis, P., & Strickland, D. (2002). Who blinks in volatile markets, individuals or institutions? *Journal of Finance*, *57*, 1923–1949.
Egginton, J. F., Van Ness, B. F., & Van Ness, R. A. (2016). Quote Stuffing. *Financial Management*, 583–608.
Gompers, P., & Metrick, A. (2001). Institutional investors and equity prices. *Quarterly Journal of Economics*, *116*, 229–260.
Harris, L. (2003). *Trading and exchanges: Market microstructure for practitioners*. Oxford: Oxford University Press.
Harris, L. (2013). What to do about high-frequency trading. *Financial Analysts Journal*, *69*, 6–9.
Hasbrouck, J., & Saar, G. (2013). Low-latency trading. *Journal of Financial Markets*, *16*, 646–679.
Hosking, P. (2006). Hedge fund returns 'are vastly overstated'. *The Times (London)*, *February*, 28.
Huang, R. D., & Stoll, H. R. (1995). *Competitive trading of NYSE listed stocks: Measurement and interpretation of trading costs.* Working Paper 94-13. Nasville, TN: Financial Markets Research Center, Owen School, Vanderbilt University.
Huang, R. D., & Stoll, H. R. (1996). Dealer versus auction markets: A paired comparison of execution costs on NASDAQ and the NYSE. *Journal of Financial Economics*, *41*, 313–357.
Investment Company Institute. (2021). *2021 Investment company factbook*. Available at <http://www.ici.org/system/files/2021-05/2021_factbook.pdf>.
Jones, C. (2013). What do we know about high-frequency trading? Unpublished Working Paper, Department of Finance, Columbia University, New York, NY.
Kim, K. (2007). *Electronic and algorithmic trading technology: The complete guide*. New York: Academic Press.
Kirilenko, A., Kyle, A. S., Samadi, M., & Tuzun, T. (2017). The flash crash: High-frequency trading in an electronic market. *Journal of Finance*, *72*, 967–998.
Kraus, A., & Stoll, H. R. (1972). Price impact of block trading on the New York stock exchange. *Journal of Finance*, *27*, 569–588.
Kwan, S. (2003, June 16). Underfunding of private pension plans. *Economic Letter*, Federal Reserve Bank of San Francisco, Number 2003-16.
Lewis, M. (2014). *Flash boys*. New York: W.W. Norton and Company.
Macey, J., & O'Hara, M. (1997). The law and economics of best execution. *Journal of Financial Intermediation*, *6*, 188–223.
McNamara, S. R. (2016). The law and ethics of high frequency trading. *Minnesota Journal of Law, Science & Technology*, *17*, 70–152.
Menkveld, A. J., Yueshen, B. Z., & Zhu, H. (2017). Shades of darkness: A pecking order of trading venues. *Journal of Financial Economics*, *124*, 503–534.
Mittal, H. (2008). Are you playing in a toxic dark pool? A guide to preventing information leakage. *Journal of Trading*, *3*, 20–33.
Patterson, S. (2014). High-speed stock traders turn to laser beams. *Wall Street Journal*.
Popper, N. (2012). Beyond Wall Street: Curbs on high-speed trades proceed. *The New York Times*.
Saar, G. (2001). Price impact asymmetry of block trades: An institutional trading explanation. *Review of Financial Studies*, *14*, 1153–1181.
Schwartz, R. A. (1988). *Equity markets: Structure, trading and performance*. New York: Harper and Row Publishers.
Shkilko, A., & Sokolov, K. (2020). Every cloud has a silver lining: Fast trading, microwave connectivity, and trading costs. *Journal of Finance*, *75*, 2899–2977.
Wah, E. (2016). How prevalent and profitable are latency arbitrage opportunities on U.S. stock exchanges? *SSRN Electronic Journal*.

4.10 EXERCISES

1. Suppose that you are a trader at a large trading firm. You have just received a sell order from a mutual fund for 1,000,000 shares of stock that has an average daily

volume of 250,000 shares. What alternatives do you have to complete this client order by the end of the trading day?

2. Suppose that Investor X executed a sale transaction at 4.22 when the stock was quoted at 4.21 bid and 4.27 ask, and that another investor's subsequent transaction executed 4.23. Calculate the quoted spread, the half spread, the effective half-spread and the realized half-spread for Investor X's execution.
3. Suppose that a client of the ABC brokerage firm seeks to have her order of $X = 500{,}000$ shares executed within three hours with the best possible execution; that is, she wants to realize the lowest possible total execution cost B. ABC works with a market impact model that estimates the cost of each of n equal-sized executions, B/n to have a fixed component F and a variable cost component $v(X_i/n)^2$, where X_i/n is the number of shares traded in any given execution i:

$$B/n = F + v(X/n)^2$$

Broker analysts have estimated F to be \$10, based on the order processing costs of each transaction. Market impact costs are estimated with the square root function with v equal to 0.0001.

   a. What are the optimal number of equal size transactions and what are the optimal transaction sizes?
   b. What are the total market impact or slippage costs associated with the 500,000 share order?
   c. Suppose that F in the market impact model is interpreted differently. Rather than representing fixed administrative costs of a transaction, $F = rn^s$, where coefficients $r = 0.1$ and $s = 0.5$ were determined as part of a new OLS regression that still determined numerical values for $v = 0.0001$ and $m = 2$, the function F is now considered to reflect the total slippage associated with n transactions for stock per hour. Thus, the market impact model is now estimated as follows:

$$B/n = F + v(X/n)^2 = rn^s + v(X/n)^2 = 0.1n^{0.5} + 0.0001(500{,}000/n)^2$$

Based on this new market impact model, what is the optimal number of equal-sized transactions and what are the optimal transaction sizes?

4. Sniffing algorithms often attempt to discern competitors' algorithms. In some instances, mere observations of competitor trading activity can reveal information as to competitor strategies, latent demand (icebergs or dark pools), and so on. Beagle Trading is a proprietary trading company that seeks to sniff out its competitors' trading algorithms in an effort to detect their likely order flows. In an effort to sniff out trade slicing models used by brokers, Beagle intends to rely on the following market impact model estimated from an OLS based on its own trading experience:

$$B/n = F + v(X/n)^2$$

where Beagle estimates the cost of each of n equal-sized executions, B/n to have a fixed component $F = 10$ and a declining variable cost component $v(X/n)^2$, where $v = 0.0001$ and X_i/n is the number of shares traded in any given execution i. Beagle observes a particular broker repeatedly placing buy orders of $c = 214$ shares of Stock GMNX in various equities marketplaces. Beagle assumes that these order sizes are optimal given its market impact model.

a. Based on the observed order sizes $c = X/n = 214$, what can Beagle infer about the latent demand for this stock? That is, is Beagle able to sniff out the total number of shares that the broker seeks to purchase based on its market impact model and observed order sizes? Why or why not?

b. Suppose that F in the market impact model is revised to be interpreted and calculated differently. Rather than representing fixed administrative costs of a transaction, F will equal rn^s, where coefficients $r = 0.1$ and $s = 0.5$ were determined as part of the same OLS that determined numerical values for $v = 0.0001$ and $m = 2$. The function F is now considered to reflect the total slippage associated with n transactions for stock per hour. Thus, this revised market impact model is now estimated as follows:

$$B/n = F + v(X/n)^2 = rn^s + v(X/n)^2 = rn^s + v(X/n)^2 = 0.1n^5 + 0.0001(X/n)^2$$

Suppose that Beagle observes that a broker has submitted orders of $c = 214$ shares. Based on these observed order sizes, and with its revised market impact model, what can Beagle infer about the latent demand for this stock? That is, is Beagle now able to sniff out the total number of shares that the broker seeks to purchase based on its market impact model and observed order sizes? Why or why not? If Beagle can infer a latent demand, how many shares does the broker intend to purchase, including transactions already observed?

5. a. Prepare an argument to support the view that the practice of payment for order flow diminishes the liquidity of markets.

b. Prepare an argument to support the view that the practice of payment for order flow enhances the liquidity of markets.

6. Critics of high-frequency trading often complain that the practice is unfair because most investors do not have the technical ability to compete with the larger, better-financed institutions. For example, some institutions pay substantial sums to co-locate trading facilities in exchange buildings to reduce latency. Exchanges have argued that co-location and HFT activities are fair because the fractions of seconds shaved off order routing and execution times are not meaningful. However, Arnuk and Saluzzi (2009) point out that "some of the exchanges make sure that each co-located customer receives equal amounts of connecting cable, so that a server at the northeast corner of a facility has the same latency as one at the southwest corner." What might this mean about the exchanges' opinions with respect to transaction speed and unfair advantages in trading?

7. a. Prepare an argument to support the view that securities markets benefit from high frequency trading (HFT).

b. Prepare an argument to support the view that securities markets are impaired by HFT.

8. Explain how and why a "speed bump" in a particular market might be beneficial to slower traders competing against HFT traders.

9. What is the difference between an "immediate or cancel" order (described in Chapter 2) and a flash order?

10. Front-running is an unethical and often illegal practice in which a broker, with the knowledge of a client's order, places her own order ahead of the client's, seeking to

profit from the client's order, perhaps affecting the market price at which the client executes, or even transacting with her own client at a profit. Describe why many flash order critics regard much flash trading to be a form of front-running.

11. How are slow traders' quotes picked off by HFT traders when the NBBO shifts?

12. What are "crumbling quotes"?

13. Pegged orders are created to follow the NBBO as it moves. In addition to the installation of the "speed bump," the IEX has provided the primary peg order, which is intended to protect small or retail traders without the fastest access to markets from traders who can react more quickly to changing market conditions. Why might such orders run afoul of the NMS and the SEC's efforts to provide traders continuous access to the NBBO?

14. When the NYSE reduced the typical minimum tick size (also known as a Minimum Price Variant—MPV) from 1/8 to 0.01, spreads and transactions costs were reduced, resulting in, according to most observers, a better quality market. Suppose that the minimum tick size were further reduced to 0.0000001 for all stocks. How or why might such a small tick size worsen markets and impede quote posting?

15. **a.** What characterizes a "fragile market"?
b. What defines a "resilient" or "robust" market?
c. Market spreads have declined considerably over the past 25 years. Why might much or at least part of this spread reduction be regarded as illusory (fake) or not really there?

16. **a.** What is meant by "naked access"?
b. How might naked access harm securities markets?
c. What is meant by "naked short sales"?
d. How might naked short sales harm securities markets?

17. Prepare an argument to dispute the following statement made by your acquaintance: "As an individual investor, I cannot compete in the stock market today. In fact, even if I had access to the best public news sources and the best analytical skills as an experienced trader, I could not compete against the extraordinarily expensive high-speed communications and execution technology employed by high frequency traders."

CHAPTER

5

Regulation of Trading and Securities Markets

5.1 REGULATORY APPROACHES AND THE REGULATORY BALANCE

The primary purpose of government regulation of competitive markets is to prevent market failure or collapse. Proponents of regulation argue that financial markets, left unregulated, will tend towards loss of competition, stability, efficiency, and credibility, causing individuals and firms to withdraw from participation. This loss of participation weakens the viability of financial markets, leading to partial or complete market failure. Senator Edmund Muskie, in his 1970 introduction of what was to become the Securities Investor Protection Act to the Senate, stated[1]:

> The economic function of the securities markets is to channel individual institutional savings to private industry and thereby contribute to the growth of capital investment. Without strong capital markets it would be difficult for our national economy to sustain continued growth: indeed, the state of U.S. capital market development, more advanced than that of any other industrial country, is an important contributing factor in the rapid economic growth this country has experienced. Securities brokers support the proper functioning of these markets by maintaining a constant flow of debt and equity instruments. The continued financial wellbeing of the economy thus depends, in part, on public willingness to entrust assets to the securities industry.

The purpose of government regulation and its enforcement is to accomplish what the market on its own cannot—establishment of a set of rules intended to maintain market competition, stability, efficiency, and credibility. Governments also use regulation to protect consumers, investors, and taxpayers and to prevent anti-trust and various illegal activities. By the mid-1930s, the United States had developed the most comprehensive securities regulatory system in the world. Other countries and the European Union have since developed regulatory systems that resemble in many respects the U.S. system.

[1]S. Rep. No. 91-1218.

Financial Trading and Investing
DOI: https://doi.org/10.1016/B978-0-323-90955-6.00005-7

© 2023 Elsevier Inc. All rights reserved.

Regulatory Powers

Financial regulators have a variety of powers to meet their goals. These powers vary among agencies and can be exercised by combinations of national, state, and local authorities. Primary regulatory powers can be categorized as follows (see Congressional Research Service [2020]):

- **Licensing, Chartering, and Registration**: Intended to ensure that financial professionals and institutions obtain and maintain appropriate credentials and suitable qualifications for conduct of their permitted activities and agree to conduct themselves in a manner consistent with regulations and practices set forth by the regulator.
- **Rulemaking**: Regulators develop and issue specific regulations, often with the assistance of members of the profession, intended to ensure fulfillment of statutory mandates.
- **Oversight and Supervision**: Regulators are empowered by governments, often with the assistance of professional associations, to monitor and ensure regulatory compliance.
- **Enforcement**: Regulators can compel regulatory compliance. Many regulators can revoke licenses, charters, and registration, administer fines and other penalties, obtain or issue cease and desist orders, and pursue criminal or civil court actions as well as administrative proceedings and arbitrations.
- **Resolution**: In some instances, regulators have the power to resolve a failing financial institution by initiating conservatorship (i.e., the regulator takes control of the firm) or receivership (i.e., the regulator liquidates the firm) or by forcing the firm into a bankruptcy process.

Regulatory Approaches

Regulatory approaches around the world can be strikingly different despite the coordination efforts of global organizations, and the extent to which regulatory structures for non-U.S. securities industries seem to resemble that of the U.S. securities regulatory authorities. Regulatory authorities tend to take some combination of two basic approaches to designing a regulatory system:

- *Rules-based approach*, where authorities set forth specific and detailed prescriptive rules to which securities markets participants must adhere. Securities regulatory authorities taking this approach often focus on process and risk, where the authority considers whether there is a potential market failure or abuse that needs to be addressed and conducts an analysis to determine how to address the problem given the constraint of limited resources. Rules-based regulation is frequently implemented as a preventive mechanism and is often associated with greater rigidity.
- *Principles-based approach*, where authorities set forth a small number of regulatory objectives and principles, focusing on results and outcomes rather than processes, and granting regulatory authorities and business firm operating and compliance officers more judgment in ensuring that policy objectives are being fulfilled.

While all regulatory authorities have used some mix of these two approaches, the United States has tended to have more of a rules-based orientation while the United

Kingdom has been the primary proponent of a principles-based orientation that is more popular with the industry. Australia (Australia Securities and Investment Commission) also employs a principles-based approach and has relatively few regulators. China employs a mix of both approaches as it moves somewhat away from a principles-based approach after its 2015 financial crisis.

Proponents of principles-based approaches argue that regulators waste less time on trivial rules-based minutiae and spend more time on meaningful deviations from appropriate practices. In addition, principles-based approaches can ease the compliance efforts of firms that engage in proper practices. On the other hand, some observers argue that principles-based regulators (such as the UK Financial Conduct Authority) tend to find very few violations and respond to violations with a light touch. This raises the question of whether few violations actually occur under this approach or are simply not discovered. Principles-based regulation might require more manpower and other regulatory resources to monitor and enforce, which might be in shortest supply when they are needed most. For example, such monitoring and enforcement problems might be most acute in a political environment where regulators are contemptuous of regulations or maintain close relationships with the firms that they monitor.

The Regulatory Balance

Running is an important activity for many people, both for health and for recreation purposes. However, dozens of people die every year from running, dozens more than die from trading securities. Yet, there are thousands of pages of federal regulation devoted to trading and none devoted to running. Why are securities markets so heavily regulated? Securities markets with coordinated mass participation are essential to the economy whereas running is satisfying either as a solo or as an uncoordinated social activity. Securities markets confer substantial rewards to those with superior information and the ability to speedily react to it. Criminals and unethical market participants are attracted to these rewards and many investors, whose capital is needed in the marketplace, are readily abused. Systemic failure of securities markets can cause economic collapse. Markets rife with fraud and adverse selection, a result of information asymmetry, raise the cost of capital to firms, drive capital and businesses to foreign markets, and restrict economic growth.

The problem of asymmetric information availability, probably the most pressing issue in securities markets, can be ameliorated with market-based measures such as reputation, monitoring, branding, bonding, independent certification, and various signaling mechanisms. However, government-based regulation by mandating truthful information disclosure, enforcing antifraud regulations and working with a civil liability system can be a crucial supplement to market-based measures, particularly when dealing with criminal and unethical market participants. Truthful information disclosure confers significant social benefits to creditors, employees and consumers, adding to the significant private benefits taken by shareholders. Fraudulent disclosure distorts business practices, distorts competition, leads to contagion and inhibits economic learning (e.g., false information disclosures by one bank detract from investors' abilities to analyze another bank).

While excessive, duplicated, overlapping, or poorly designed regulation impairs markets, all regulation, even well-designed regulation, is expensive (see Section 5.8 on regulatory costs). We will discuss in greater detail the costs, both direct and indirect, of securities regulation later in this chapter in Section 5.8, as well as in Chapter 13, Trading Gone Awry.

5.2 PRE-1930s SECURITIES REGULATION: THE BACKGROUND

Securities market regulation, with some notable exceptions, is primarily a fairly recent phenomenon. Securities market abuses were, for the most part, limited by more general statutes, common law and court decisions (e.g., *Strong v. Repide, 1909*, discussed later in Sections 5.4 and 13.2) related to fraud and other offenses. However, there were a few earlier attempts to enact certain securities market regulations, and these earlier efforts provided the framework for more modern regulation. For example, under King Edward I in 1285, broker licensure was instituted after the finding that English brokers provided less satisfactory service than their Italian counterparts. Active securities trading in London late in the 17th century accompanied by a monetary crisis in 1695 led to the *Brokers Act of 1696*, which required stock brokers to be licensed and limited the number of these licenses to 100 for approximately 10 years. Following the South Sea Bubble, but before the South Sea Company's crash, the *1720 Bubble Act* passed in Britain only shortly after the creation of what was to become the London Stock Exchange (See Section 2.2). The Act, which the South Sea Company hoped would preserve its monopoly, provided for the issuance of security prospectuses and prohibited certain types of trading fraud, but was ultimately repealed in 1825 (See Mahoney, 2015). The *Companies Act of 1844* provided once again for company issuances of prospectuses, and the *Companies Act of 1867* along with subsequent legislation provided for specific details in those prospectuses. This 19th-century British legislation, based largely on full disclosure of material information, served as an important conceptual foundation to 1930s U.S. securities legislation.

Securities legislation activity in the United States was very slow to start. Massachusetts enacted one of the earliest securities regulations in 1852, requiring railroad companies to certify that "responsible parties" had subscribed to their stock. The Constitution of the State of California signed in 1879 prohibited the sale of stock on margin. At the federal level, Congress made a small number of early attempts to regulate securities trading. For example, it passed and then quickly repealed the *Anti-Gold Futures Act of 1864*, intending to restrict trading in gold and exchange contracts. Congress attempted to regulate agricultural futures trading with the *Future Trading Act of 1921*, but this Act was found by the U.S. Supreme Court to be unconstitutional.

The Blue Sky Laws

In the early 1900s, several states passed limited legislation to regulate securities markets. These earlier laws sought to prevent issuance of securities that were considered to be unfair or did not promise a "fair return." Securities were potentially subject to merit tests. For example, in response to numerous incidents of brokers having sold worthless

securities of sham companies, Kansas enacted a securities law in 1911 requiring registration of brokers and securities. These worthless securities were said to be backed by "nothing but the blue skies of Kansas." This state legislation was considered to be the first of state *blue sky laws* intended to regulate securities markets.[2]

Prior to the 1930s, most regulation of U.S. securities markets was provided by these so-called blue sky laws, which critics often regarded as permitting "anything under the blue sky." The *Uniform Securities Act* was adopted by a number of states as early as 1930 in an effort to coordinate the state-based securities legislation. The Act has been revised extensively over the years, but has not been adopted by all states, including New York.

New York State Regulation and Enforcement

While most early blue sky laws tended to be rather ineffective against many securities markets abuses, a small number of state securities laws can be applied effectively to combat a variety of financial wrongdoings, even when the U.S. Justice Department and the SEC cannot or will not pursue corrective action. In particular, New York State's Martin Act, enacted in 1921, amended several times and creatively applied by several New York State attorneys general, notably Albert Ottinger and Elliott Spitzer, has been used to collect billions of dollars in fines and settlements and to force security firms to modify fraudulent and questionable practices. For example, in 2001, Spitzer used the extraordinary investigatory powers conferred by the Martin Act to launch an investigation into conflicted investment advice provided by Merrill Lynch's star technology analyst, Henry Blodgett, who hyped stocks marketed by Merrill's investment banking department to investors while trashing them in private emails. Merrill settled by paying $100 million in fines and changing its method for compensating stock analysts, after seeing its stock price drop by $5 billion in a single week after Spitzer publicly leveled his accusations. Similar actions based on similar accusations were taken against Salomon Smith Barney and 9 other investment banks yielding a combined total of $1.4 billion in fines. More recently, New York State Attorney General Eric Schneiderman alleged that the dark pool of Barclays bank, Barclays LX, had "falsified marketing material purporting to show the extent and type of high frequency activity in its dark pool," and ultimately extracted settlements of $154 million from Barclays and Credit Suisse Securities. Some critics argue that the Martin Act has been abused by New York attorneys general to extort excess payments from financial institutions and to generate publicity to promote their efforts to seek higher office (e.g., Ottinger, Jacob Javits, Spitzer, and Andrew Cuomo have been so accused).

How does New York's Martin Act enhance the attorney general's investigatory powers and prosecutorial leverage? First, the Act applies to all fraudulent securities transactions that occur, have victims or are perpetrated by firms with a presence in New York. This applies to nearly all large financial institutions. The Martin Act gives wide subpoena and evidence-gathering power to attorneys general, even in the absence of probable cause or a Grand Jury. Prosecutors have the option to release relevant documents to the public,

[2]The precise meaning and origin of the term "Blue Sky Laws" are somewhat unclear. See Fleming (2011). See Loss and Cowett (1958) for more details on "Blue Sky Laws."

risking reputational damage to accused firms and individuals. This broad investigatory power gives the attorney general significant tactical advantages, encouraging defendant cooperation. Unlike those in federal fraud cases, convictions under the Martin Act do not require proof of intent to defraud or even that fraud victims suffered monetary losses, only that the accused "made a material misrepresentation or omission during a securities transaction" (Sgambati, 2016).

Market Abuses: An Introduction

However, in recent years, and despite many criticisms and shortcomings, some observers consider the United States to have the most effective national securities regulatory system in the world. Securities market regulation and enforcement are currently provided by a web of legislation, regulatory agencies, and professional associations. This regulation is largely intended to manage systemic risk, prevent price manipulation, and protect the investing public against securities market abuses. The following are just a few of the many types of activities and abuses such regulation is intended to restrict[3]:

- *Wash sales* are transactions intended to create the appearance of sales where, in effect, no sales actually take place. Wash sales may be intended to manipulate security prices (e.g., conspirators execute transactions with one another to create records of sales prices to deceive other participants in the market). Wash sales may also involve the sale of securities for tax purposes with offsetting transactions to repurchase the securities or related instruments.
- *Corners* involve the purchase a sufficient level of a given security to obtain market power over its price.
- *Churning* occurs when a broker executes an inordinately large number of transactions on a client's account for the purpose of securing commissions.
- *Pools* exist when market participants combine their resources to obtain a large position in a security in order to manipulate its price.
- *Fraud* constitutes an intentional deception or misrepresentation for the purpose of securing a profit.
- *Inside trading* constitutes trading on the basis of material nonpublic information.

These and other market abuses might have been among the causes of the stock market crash of 1929. On September 3, 1929, the Dow Jones Industrial Average (DJIA) had reached a high of 381; by July 8, 1932, it had fallen to a low of 41. Numerous other economic and structural problems have been associated with this crash including massive unemployment, deflation, failure of over 5000 banks, and so on. Market abuses of the 1920s have often been blamed for contributing to the Great Crash and perhaps to the Great Depression as well (the causes of the Great Depression remain subject to much controversy). Nevertheless, the first major federal efforts at curbing market abuses resulted from these tumultuous events.

[3]There will be more discussion of securities markets abuses in Chapter 13.

5.3 U.S. SECURITIES MARKET LEGISLATION: THE FOUNDATION

Beginning in the 1930s, a series of regulatory acts were proposed to prevent or mitigate market failures such as the Great Crash of 1929. The financial system was in shambles due to the devastated economy, and there was a clear need to restore integrity to financial markets to rebuild the economy. Government involvement was clearly needed to develop a much-needed regulatory system. The following sections discuss some of the major U.S. legislative acts intended to restrict abuses such as those discussed above and to promote fairness and transparency in the securities markets. Such sweeping legislation was made possible, in part, due to overwhelming Democrat majorities elected to both houses of Congress and the election of President Roosevelt in 1932. Twenty-five days after his inauguration in 1933, President Roosevelt asked Congress for a new law that would "put the burden of telling the whole truth on the seller" of securities, and, referring to the *caveat emptor* rule generally preferred in business circles, added: "Let the seller also beware."[4] The financial community was in a poor position to effectively protest this imposition of regulation at the height of the Great Depression, though many members did argue that the Act would be burdensome, stifle entrepreneurship, drive business offshore and make independent directors reluctant to sit on corporate boards. Some institutions even threatened to protest the Act by refusing to bring new issues of stock to the market in a "Wall Street strike."

The Bedrock: The Securities Act of 1933 and the Securities Exchange Act of 1934

The Securities Act of 1933, sometimes called the "Truth in Securities Law," deals primarily with new issues of securities. The Act requires that issuers and underwriters provide financial and other significant information concerning securities offered for public sale. In addition, the Act prohibits deceit, misrepresentations, and other fraud in the sale of securities. Unlike most of the "blue skies laws" that focused on the merits of securities, the Securities Act focused on making full, accurate and timely information available to prospective investors in securities. Its major provisions are as follows:

- All primary issues must be registered with an appropriate government agency (later to be the Securities Exchange Commission or SEC). The registration will include proper statements and documentation.
- A prospectus must accompany each new issue. This prospectus must contain a complete and accurate accounting of the firm's condition, risks, and prospects, and state how the proceeds of the new issue will be used.
- Small and private issues are exempt from the registration provisions. In addition, SEC Rule 415 (shelf registration) of 1982 allows up two years for securities to actually be issued after completing the SEC registration process.
- Firms, officers of firms, and underwriters are prohibited from making false statements regarding their new issues, and may be criminally liable for doing so.

[4]See Schlesinger (1958), p. 441.

The Securities Exchange Act of 1934 was primarily intended to improve information availability and to prevent price manipulation. While it was signed into law shortly after the Securities Act of 1933, it was brought forth by Congress at roughly the same time and was intended to complement the Securities Act to form the framework for the new U.S. securities regulatory system. Whereas the Securities Act of 1933 dealt mainly with primary issues, the Securities Exchange Act of 1934 dealt mainly with secondary markets, accomplishing the following:

- Established the SEC as the nation's primary federal securities regulatory authority.
- Provided for annual and other periodic reporting by public companies.
- Limited insider trading activity.
- Provided rules for proxy solicitation.
- Required registration of exchanges.
- Provided for credit regulation: Regulations T (for brokerage firms) and U (for nonbroker lenders) permit the Board of Governors of the Federal Reserve to set margin requirements. Since 1974, investors have been required to post 50% margin (deposit or collateral) when purchasing stock on margin. FINRA Rule 2520 permits pattern day traders 25% initial margin requirements (subject to a $25,000 account equity balance).
- Subjected institutions to the Net Capital Rule, amended in 1975, imposing limits on broker-dealer debt-to-net capital ratios. Certain exceptions were made in 2004 to five of the largest institutions, all of which were bankrupted, merged, or otherwise suffered significant financial distress in the financial market crisis of 2008.
- Prohibited securities price manipulation.

Additional Major Depression-Era Legislation

During the 1920s, banks were heavily engaged in trading, investments in riskier securities, and underwriting of securities. Higher risk nonbanking activities led to bank failures. The *Banking Acts of 1933 (Glass Steagall)* was motivated by the failure of thousands of banks.[5] Glass Steagall was intended to help restore faith in the banking system and prevent a similar crisis from reoccurring. This law fundamentally altered the U.S. banking system. There were three very important features of the legislation. First, the Federal Deposit Insurance Corporation (FDIC) was created to provide federal insurance on bank deposits (subject to a current $250,000 ceiling). The second provision (and most relevant to trading) of the Glass-Steagall Act was the imposition of restrictions on the activities of commercial banks. It prevented commercial banks from underwriting securities, trading with the public (excepting certain U.S. government and municipal bonds and real estate loans), and owning risky securities such as corporate stocks and bonds. These regulations separating financial functions were largely in response to the fact that many banks in 1933 had lost much of depositors' money to trading securities and in their securities underwriting services. Much of this provision was relaxed by various court decisions,

[5]There was an earlier first Glass-Steagall Act, enacted in 1932 in reaction to deflation, dealing with the federal government's ability to trade debt securities. This act had a much smaller effect on the U.S. banking system.

regulator policy, and the Gramm-Leach-Bliley Act of 1999, with significant consequences.[6] The third important provision of the Glass-Steagall Act placed various regulations on the interest rates banks were allowed to pay (Reg Q), although most of these provisions were relaxed in the 1980s.

The Commodity Exchange Act of 1936 provided for regulation of commodities and futures trading markets by the Department of Agriculture and required all futures and commodity options to be traded on organized exchanges. Ultimately, the Act led to the 1974 establishment of the Commodities Futures Trading Commission (CFTC).

The Maloney Act of 1938 was essentially an amendment to Section 15.A of the Securities Exchange Act of 1934. It provided for the self-regulation of the over-the-counter (OTC) markets, leading to the establishment of the National Association of Securities Dealers (NASD, which merged into FINRA approximately 70 years later). It provided that the SEC can overrule the NASD and provided for NASD penalties. The Act has been hailed as "an especially provocative exercise of governmental power by a private organization" (see Hed-Hofmann, 1965).

The Trust Indenture Act of 1939 provided that issuers of bonds must clearly specify rights of their purchasers in an agreement known as a *trust indenture*, provide bond trustees with periodic financial reports, and that the trustee not impair purchasers' right or willingness to sue the issuing corporation. Essentially, this Act extended parts of the Securities Act of 1933 to primary bond markets.

The Investment Companies Act of 1940 extended the Securities Act and the Securities and Exchange Act to investment companies such as mutual funds, which were not prevalent prior to the 1930s. This Act provided that investment companies must:

1. avoid fraudulent practices.
2. fully disclose financial statements.
3. distribute prospectuses.
4. publish statements outlining goals, which will not be changed without appropriate process.
5. limit issuance of debt.
6. follow uniform accounting procedures.
7. operate the fund to benefit owners, not managers.

The Investment Advisors Act of 1940 requires advisors and certain investment institutions who provide clients with paid investment advice to register with the SEC. The Act focused more on NASD (now FINRA) and the SEC being able to identify those acting as investment advisors than on regulating and restricting their activities. However, the Act did specify certain illegal advertising practices by advisors. The Investment Advisors Act also established the *fiduciary standard*, which holds that a *Registered Investment Advisor* must put her client's interests above her own, following the very best course of action, regardless of how it affects the advisor personally. That is, the advisor's primary loyalty is to the client. The *suitability standard* is less strict than the fiduciary standard, with the suitability standard holding that advisors must provide clients with suitable advice, appropriate to the client's needs, but not necessarily the best advice, as long as it is suitable (see FINRA Rule 2111).

[6]See Markham (2010).

In contrast to the fiduciary standard applicable to the registered investment advisor, the advisor's first loyalty under the suitability standard is to her employer or to herself. The Act was amended in 1996 to require only those advisors with over $25 million under management to register, although other advisors may register as well.

5.4 CRISES AND UPDATING THE REGULATORY SYSTEM

As in the 1930s, most subsequent securities market legislation was born of securities market booms and resulting crashes. This regulatory development process has often been criticized as "fighting the last war," in that much legislation seems intended to prevent the most recently occurring crisis rather than the next potential crisis.

The 1960s Go-Go Years and Reform

The securities brokerage industry underwent another crisis in the late 1960s, with many firms failing due to their inability to handle their vastly increased volume of paperwork. Brokers during the paperwork crisis struggled with their unprecedented success as the average NYSE daily trading volume increased from 3,042,000 shares in 1960 to 12,971,486 in 1968. Major firms such as Goodbody & Co., Pickard & Co., and F.I. Dupont, Glore Forgan & Co. either failed or had to be rescued. Hundreds of millions of dollars in stock certificates vanished, at least some due to organized crime, while billions of dollars in transactions failed. Senator Edmund Muskie introduced legislation to address this market failure and to restore investor confidence in markets. The *Securities Investor Protection Act of 1970* was passed, and provided for the establishment of the *Security Investors Protection Corporation* (SIPC) to insure accounts of brokerage clients. SIPC is funded with brokerage firm dues and protects clients from losses due to brokerage firm failure. The range of coverage provided by SIPC insurance should not be overestimated because it does not protect clients from trading losses or even most types of brokerage fraud. Instead, SIPC replaces missing cash, shares of stock, and other securities in the event of brokerage firm failure. SIPC generally does not provide for protection of commodity futures contracts, investment contracts (such as limited partnerships), and unregistered fixed annuity contracts.

When the auto company Studebaker Corporation failed in 1963, its pension plan was unable to fulfill its obligations. Senator Jacob Javitz introduced legislation to address failing employee pension and health plans. The eventual result was the *Employee Retirement Security Act of 1974* (ERISA), which makes provisions for preventing pension funds abuse and prohibits imprudent risks taken by fund managers. The Act provided for the use of the *prudent man rule*, but left it to the court system to clearly define it. Generally, the prudent man rule holds that a fiduciary is required to invest trust assets as a "prudent man" would invest his own property with the following factors in mind: the needs of beneficiaries, the need to preserve the estate (or corpus of the trust), and the amount and regularity of income. ERISA broadened the scope of this rule without making clear specifications, while holding fund managers personally responsible for "imprudence losses." In addition, ERISA made provisions for vesting rules and established the Pension Benefit Guarantee Corporation.

Updating Commodities Trading Regulation

The *Commodity Futures Trading Commission Act of 1974* created the Commodity Futures Trading Commission (CFTC) and transferred regulatory authority over futures markets from the Department of Agriculture to the CFTC. The Act also gave the CFTC regulatory authority over nonagricultural futures and options on those futures. In 1982, the CFTC provided for the creation of the National Futures Association (NFA), a self-regulatory body for commodities and futures traders.

The National Market System

The brokerage crises were still worrisome to markets and legislators, and a more manageable clearance and settlement system was needed to avoid another paperwork crunch. Furthermore, many legislators were concerned about the fixed commission structure maintained by brokerage firms. As discussed in Chapter 2, The *Securities Acts Amendments of 1975* abolished this long-standing practice of fixed brokerage commissions and called for the establishment of a "national market system" (NMS) where orders would be routed to the best market. This led to efforts to establish a single CLOB (consolidated limit order book) that would list all limit orders and dealer quotes, and resulted in the continuous calculation and dissemination of the NBBO of all listed stocks to all broker-dealers along with the requirement that they pursue best execution (and later, best price) on behalf of their customers.

The national market system has led to enhanced competition in which brokerage firms, dealers, and markets would compete with each other in a national market intended to be unhindered with unnecessary regulation. In addition, the amendments, implemented on May 1, 1975, sometimes known as "May Day" in securities markets, have led to a national system for clearing and settlement. These amendments have led to the Securities Information Processor (SIP; See Section 3.2), the creation of the consolidated trade system (CTS) and the consolidated quote system (CQS) or more generally, a national market system.

The national market system expanded competition among trading arenas as we will detail in Sections 5.5 and 5.8. The *Unlisted Trading Privileges Act of 1994* allowed for listed stocks to be traded in any venue, creating distinctions between listing (primary) and trading exchanges for various stocks. A company meeting the requirements of an exchange can have its shares qualify for primary listing on that exchange. After its primary listing, unlisted trading privileges (UTPs) allow the company to then apply to the SEC to have its shares traded on other U.S. exchanges (trading exchanges) as well. Thus, for example, Microsoft, a Nasdaq-listed stock (Nasdaq is Microsoft's primary exchange), can also trade on the NYSE. UTPs can enhance competitive trading opportunities for such a stock and provide fees to trading exchanges for improving quotes, although UTP does complicate matters for the SIP (see Section 3.2), which must collect and aggregate bid, offer, and sales data from all exchanges and compile the NBBO and quoted spread for each stock.

Insider Trading Regulation

Insider trading violations were infrequently pursued by U.S. authorities until the 1980s, when numerous highly profitable incidents generated significant publicity. But what is

insider trading? When is insider trading illegal? First, illegal insider trading is traditionally characterized as the execution of securities transactions on the basis of material nonpublic information, in the U.S., in violation of a fiduciary trust. However, there is no comprehensive statutory definition in the U.S. as to precisely what constitutes impermissible insider trading or even what constitutes an insider. Regulators generally rely on several statutes to enforce prohibitions on insider trading. First, Section 14(e)-3 of the Securities Exchange Act essentially bans trading in securities of a tender offer target by anyone (other than the bidder) on the basis of material, nonpublic information relating to the tender offer. The Act extends to those acting on behalf of either the target or bidder. Insider trading by officers, directors, and direct or indirect beneficial owners of more than 10% of equity is expressly prohibited in Section 16 of the Act. Section 10(b) of the Act, which prohibits the employment "in connection with the purchase or sale of any security registered on a national securities exchange or any security not so registered, any manipulative or deceptive device or contrivance in contravention of such rules and regulations as the Commission may prescribe as necessary or appropriate in the public interest or for the protection of investors." SEC Rule 10b-5 prohibits "any act, practice, or course of business which operates or would operate as a fraud or deceit upon any person, in connection with the purchase or sale of any security." Together, Section 10(b) and Rule 10b-5 prohibit making false statements or omitting material facts related to the purchase or sale of securities. However, while neither specifically mentions insider trading, broad court interpretations of the word "fraud" in Rule 10b-5 include breach of the personal or fiduciary obligation of an insider to maintain confidentiality, and extends to certain "tippees" of those who fail to fulfill their obligations.

Subsequent to several insider trading scandals in the early 1980s, Congress passed the *Insider Trading Sanctions Act of 1984*, which authorized penalties for illegal insider trading equal to three times the illegally obtained profits plus forfeiture of the profits. However, more scandals followed. The *Insider Trading and Fraud Act of 1988* was intended to help define exactly what constitutes an insider and to increase penalties for illegal insider trading activity. This Act increased the maximum imprisonment from 5 to 10 years, with maximum criminal fines increased from $100,000 to $1,000,000 for individuals and from $500,000 to $2,500,000 for nonindividuals. The Act also initiated a bounty program allowing the SEC to pay rewards to insider trading informants.

Members of the U.S. Congress were exempt from the provisions of this legislation, which facilitated repeated profitable incidents of congressional insider trading. The 2012 passage of the Stop Trading on Congressional Knowledge (STOCK) Act finally attempted to restrict the insider trading activity and "tippee" potential of members of the legislative, executive, and judicial branches of government along with their staffs, at least with respect to inside information obtained through the conduct of their official duties. Nevertheless, as we will discuss later, violations and potential violations have continued to occur.

Supreme Court Interpretations

Several U.S. Supreme Court and other federal court cases have also played key roles in determining what constitutes illegal insider activity. An early example was the 1909 case

of *Strong v. Repide*, introduced above in Section 5.2, in which the Supreme Court established that under certain special circumstances, a company official must reveal his identity and nonpublic information when he trades with shareholders.

The 1980 Supreme Court decision in *Chiarella v. United States* narrowed the definition of a corporate insider to be one who maintains a relationship of trust and confidence with shareholders. This definition has been broadly interpreted by the SEC. For example, the SEC pursued a case against the analyst Raymond Dirks who tipped some of his firm's clients about a fraud that he uncovered involving an insurance company, which enabled his clients to secure trading profits on the basis of material non-public information. However, Dirks did not breach his relationship of trust and confidence with his clients; his tippers, who had directed him to an ongoing fraud in a whistle-blowing action, had not breached their fiduciary duty, as they had asked Dirks to take the information public, which he ultimately did. In its 1983 decision of *Dirks v. SEC*, the Supreme Court held that insiders violate the law when they knowingly act as tippers while receiving personal benefit from the tippee, which Dirks had not (See Macey, 2017).

Todd Newman, a trader at the hedge fund Diamondback Capital, had made $72 million trading Dell stock based on earnings information provided by a tippie who obtained it from a casual friend, who was a Dell insider. Thus, Newman was a "downstream" tippie. In its 2014 decision of *United States v Newman*, the Supreme Court narrowed further its Dirks decision along two dimensions. First, the personal benefit received by the tipper must be consequential, not merely reputational, "career advice" or enhancement of a friendship. Second, the tippie must be aware of this personal benefit. This limits the ability of the government to pursue downstream tippies such as Newman. Essentially, the Supreme Court determined that tippers who knowingly received consequential personal benefit from their tips broke the law, while those who received no benefit did not. In its 2016 decision of *United States v Salmon*, the Supreme Court extended the personal benefit to include gifts of inside information to personal friends and relatives, which had been noted earlier in the 1983 Dirks decision.

A primary difference between the E.U. 1989 insider trading directive (updated and strengthened by the Market Abuse Regulation) and U.S. insider trading law is that in the E.U., tippees receiving inside information from tippers who cannot trade are also prohibited from trading. In the U.S., the *Dirks* case established a personal benefit test. After Newman, the tipper is sanctioned when she receives a consequential personal benefit, and when this personal benefit is known to the tippee. In effect, the E.U. prohibits trading on the basis of inside information by anyone who knows or ought to have known that the basis of trading was inside information. Unlike in the U.S., E.U. securities regulators can pursue insider trading sanctions regardless of whether there was a violation of a fiduciary duty or obligation.

Presumably indicative of motive, this "personal benefit test" imposed by the U.S. Supreme Court has created difficulties for the SEC. Seeking to "level the playing field" and maintain the integrity of markets, the SEC has historically viewed exploiting informational advantages via inside trading as being harmful to market integrity ("fairness" approach). However, the Supreme Court's rulings suggest that it views all information-based trading as reflecting informational advantages, and seeks instead to limit illegality of insider trading to breaches of fiduciary trust and confidence against firms ("property

rights" or "misappropriation" approach). That is, if inside information is property that belongs to shareholders, insiders cannot misappropriate this information and legally trade on this inside information.

5.5 DEREGULATION, CORPORATE SCANDALS, AND THE FINANCIAL CRISIS OF 2008

After over 40 years of increasingly burdensome regulation in the financial services industries, President Ronald Reagan sought to reduce this regulatory burden, most notably in the banking industries. A near-collapse of the savings and loans industry shortly followed. Wendy Gramm (spouse of Senator Phil Gramm, an important proponent of deregulation), as chair of the CFTC, in 1989 and 1993 exempted a number of swaps and derivative instruments from regulation. These exemptions were broadened in 2000 by the Gramm-Leach-Bliley Act. The Federal Reserve Board reinterpreted the Glass-Steagall Act several times during the 1980s and 1990s so as to ultimately permit bank holding companies to earn up to a quarter of their revenues from investment banking activities.

Deregulation

The *Financial Modernization Act of 1999*, also known as the Gramm-Leach-Bliley Act, contributed to the consolidation of the financial services industries, allowing for the formation of "mega-banks." This act formally permitted commercial banks, investment banks, and insurance companies to consolidate, repealing the most important provisions of the Glass-Steagall Act passed in the 1930s to separate commercial banking from underwriting and other financial activities. While many combinations had already occurred before passage of the Act (usually through subsidiaries), their legality was questionable. The pending combination of Citigroup (commercial banking) and Travelers (insurance, investment banking, and stock brokerage) into the world's largest financial institution accelerated and contributed to passage of this Act. Mergers that would have been impossible prior to its passage have masked severe operating and financial difficulties, simply forestalling some inevitable failures. But more importantly, this Act was not replaced with or accompanied by significant regulatory oversight.

The Commodity Futures Modernization Act of 2000 exempted most OTC nonagricultural derivatives and transactions between "sophisticated parties" from regulation under the Commodity Exchange Act (CEA) or as "securities" under other federal securities laws. This act excluded most OTC energy trades from CFTC oversight and financial derivatives from SEC and CFTC oversight. These exemptions formed the so-called "Enron loophole" that contributed to massive fraud and the failure of the Enron company in addition to the role that credit default swaps would play in the 2008 financial crisis.[7] In part, the Act was intended to resolve disputes between the SEC and CFTC concerning overlapping jurisdictions, particularly

[7]See Sections § 2(g) and (h) of the Commodity Futures Modernization Act of 2000, and Cravath, Swaine and Moore (2001).

with respect to certain types of contracts including *single equity futures* (futures contracts on shares of a single firm's stock). The Act led to retail trading of these contracts in 2003.

Sarbanes-Oxley

The Sarbanes-Oxley Act of 2002 (SOX) was enacted after a wave of corporate scandals (including Enron, WorldCom, and Global Crossing) in the late 1990s and early 2000s. The Sarbanes-Oxley Act, also known as the *Corporate and Criminal Fraud Accountability Act*, was passed to provide for accounting reform, improved financial reporting, reduced conflicts of interest, and increased penalties for securities fraud. The Act contained 11 sections. Among the highlights of this Act are:

- the creation of a five-member *Public Company Accounting Oversight Board* (PCAOB) to oversee public auditing firm practices (Section I)
- the requirement for CEOs and CFOs to personally certify the accuracy of their firms' periodic reports
- restricting public accounting firms from providing nonauditing services contemporaneously with auditing services to prevent certain conflicts of interest (Section II)
- limiting company loans to directors and officers (Subsection 402)
- prohibitions of share trading by officers and directors during certain "blackout" periods
- the requirement that attorneys representing reporting companies report material violations of law or breaches of contracts to appropriate corporate authorities
- Subsection 404 of the Act, which requires that companies prepare assessments of their internal controlling practices along with auditors' reports on those assessments
- requirement that firms to disclose material off-balance sheet transactions and relationships
- prohibitions on unfair allocations of securities in the initial public offering (IPO) process
- prohibitions on misleading research reports issued by financial advisors and advisory services
- requirements for publicly traded firms to have independent directors with financial expertise serve on their audit committees

Post-SOX

Another crucial deregulatory action occurred in 2004 when the SEC relaxed the "net capital rule" (Rule 15c3-1) for the five largest investment banks. This enabled the largest investment banks to substantially increase their debt, both increasing their risk of failure and enabling them to invest more heavily in riskier assets. The rule was originally imposed in 1975 (updated from 1934), requiring that brokers and dealers doing business with the public value all of their tradable assets at market prices, discounting them to account for their market risk. In many respects, relaxation of this Act was consistent with the spirit with the Basel II Accord. Riskier assets were to be discounted more deeply than safer assets. The rule requires broker dealers to limit their debt-to-net capital ratios to 12:1. Broker dealers qualifying for exemptions under this 2004 relaxation were to be designated

as "consolidated supervised entities" (CSEs). Goldman Sachs, Morgan Stanley, Merrill Lynch, Lehman Brothers, and Bear Stearns qualified as CSEs, enabling them to maintain debt-to-equity ratios exceeding 30:1. Four of these five firms failed or were bailed out between 2007 and 2009.

The Securities Acts Amendments of 1975 led to creation of the National Market System (see Section 5.4). The Unlisted Trading Privileges Act of 1994 allowed for stocks listed on any exchange and certain OTC stocks to be traded in any venue. Thus, for example, Unlisted Trading Privileges allows companies that conduct an IPO and list on the NYSE to have their stock traded in other venues such as Nasdaq, Citadel, and Bats. Regulation NMS, which followed up on these legislative actions was adopted by the SEC in 2005. Reg NMS was intended to modernize the regulatory structure of U.S. equity markets, reflecting the accelerating pace of change in markets due to improved technology and the increased variety of types of market centers (exchanges, crossing networks, ECNs, dark pools, etc.). The key and most controversial provision of this regulation is the Order Protection Rule (Rule 611, also known as the "trade-through" rule). This rule requires that exchanges, alternative trading systems, and OTC market makers maintain and enforce policies to ensure the execution of trades at their best prices (no worse than the NBBO), and encourages traders to post aggressive limit orders, knowing that their quotes will not be bypassed for worse quotes. Regulation NMS was intended to enhance client order protection and to improve client transaction executions by facilitating intermarket access. The regulation sought to reduce spreads by allowing for subpenny pricing for sub-$1 shares and to improve public access to market data. Finally, Rule 610(c) of Reg NMS caps fees for access to quotations to $0.003 per share. In effect, this fee is a liquidity-taker fee applying when a market order is executed against the limit order associated with the quote.

The Credit Rating Agency Reform Act of 2006 was enacted to establish guidelines for nationally recognized statistical ratings organization (NRSRO) qualification, to improve regulatory oversight in the ratings industry and to reduce certain conflicts of interest and industry abuses. Prior to enactment of this Act, only Standard & Poor's, Moody's Investors Service, Fitch Ratings (which, together controlled 95% of the market), and A.M. Best and Dominion (now, DBRS) were recognized as NRSROs. The Act sought to curb the practice of sending a company unsolicited ratings along with a bill. The Act prohibited packaging ratings with the purchase of consulting and other services. For example, the credit rating agencies had been taking fees for rating derivative instruments such as mortgage-backed securities (MBSs) and collateralized debt obligations (CDOs) after having received payments from their issuers for helping to package them. Thus, agencies would advise on structuring packages, and then rate the packages that they helped structure so as to receive their highest credit ratings. The Act was intended to limit conflicts of interest that arose when MBSs, CDOs, and other securities are packaged and rated by major rating agencies. In addition, the Act required credit rating agencies to maintain policies and procedures to be in place to manage these types of conflicts when they arose.

The 2008 Financial Crisis

In the summer of 2007, there were increasing reports of troubled mortgages and weakening of the securitized assets and portfolios that contained them. These reports followed

on the peak of U.S. housing prices. By the end of the year, Countrywide Financial Corporation, one of the largest banks in the United States, was seeking shelter and Northern Rock in the UK was on the brink of failure. The financial system took a significant blow in March 2008 when Bear Stearns, one of the largest investment banks in the United States, collapsed, to be rescued by JP Morgan Chase with backing from the Federal Reserve. The special lending facilities of the Federal Reserve opened the discount window to investment banks to prevent an industry-wide collapse. IndyMac Bank FSB, another major depository institution, failed in July 2008. However, the crisis moved into full swing in September 2008 when Freddie Mac and Fannie Mae were placed under U.S. government conservatorship, Lehman Brothers filed for Chapter 11 bankruptcy protection, and the United States entered its worst recession since the Great Depression.

The contribution of trading activities to the 2008 financial crisis continues to be debated, as we discuss in the next section. Clearly, many institutions with the most substantial trading activities were among those that failed or required bail-outs. However, many of their problems might have been less a result of trading and more a result of excessive origination of toxic instruments, excess leverage, and poor regulatory oversight.

5.6 DODD–FRANK

Perhaps the most significant piece of securities legislation passed since the 1930s was the *Dodd–Frank Wall Street Reform and Consumer Protection Act* of 2010. Responding to the financial crisis of 2007–2009, Congress passed major reform covering and extending beyond securities markets intended to:

> promote the financial stability of the United States by improving accountability and transparency in the financial system, to end "too big to fail," to protect the American taxpayer by ending bailouts, [and] to protect consumers from abusive financial services practices.[8]

This 848-page act was intended to promote financial stability and consumer protection, and extended well beyond securities trading. Among the important reforms affecting traders are:

- Providing for the establishment of the *Financial Stability Oversight Council* (FSOC) to supervise systemic risk, promote market discipline, and respond to threats to financial stability. The Council is chaired by the U.S. Treasury Secretary and is comprised of 10 voting members and 5 nonvoting members.
- Providing new rules for transparency, independence, and accountability for credit rating agencies, and creation of the Office of Credit Ratings.
- Providing for the so-called *Volcker Rule*, regulating and limiting banks, their affiliates, and nonbank financial institutions supervised by the Fed with respect to proprietary trading and investment in and sponsorship of hedge funds and private equity funds.

[8]Dodd–Frank Wall Street Reform and Consumer Protection Act.

- Authorizing the Federal Reserve Board or FSOC to designate clearing agents as systemically important and to oversee their activities.
- Providing for regulatory authority over swap agreements between the SEC (security-based swaps) and the CFTC (all other swaps).
- Rather than mandate that all swap contracts be traded on exchanges, Dodd–Frank provided for the creation of a swap execution facility (SEF), specifically designed to provide for trade transparency, encourage competitive execution, and ensure a complete record and audit trail of trades, all designed to enhance swaps markets.
- Requiring companies selling credit and mortgage-backed products to retain at least 5% of the instruments' credit risk unless the underlying loans meet certain standards that reduce risk.
- Prohibiting disruptive trading practices, specifically banning spoofing by name in futures and other derivatives markets (See Section 12.4 of this text).

The Volcker Rule (so-named because of its chief proponent, Paul Volcker, former Chair of the Federal Reserve) component of Dodd–Frank restricts proprietary trading by commercial banks, with certain exceptions. The rule's primary justifications are that proprietary trading is inherently speculative and exploits essential protections and supports granted by the government to banks, including deposit insurance, routine access to Fed credit, and emergency support (See Volcker, 2012). Stability in the commercial banking system merits such support because of banking's essential roles in the overall economy including its centrality to the payments system, extension of credit, and safe-keeping of liquid assets. Proprietary trading provides little direct support for any of these key banking functions. Furthermore, proprietary trading distracts banks from performing their core functions, deposit-taking and relationship lending. Critics of the Rule object that it is complicated and costly, adversely affects the competitiveness of U.S. banks, imperils liquidity in financial markets and that proprietary trading is neither a significant contributor to overall bank revenues or significant bank risk factor. We further discuss these and other costs in Section 5.8. Proprietary trading's role in the 2008 financial crisis is still being debated, though most observers agree that its role was not central, including Volcker himself (Whitehead, 2011). Instead, most observers attribute more blame for the crisis to credit (e.g., mortgages) and liquidity risk (e.g., repos) rather than to market risk and proprietary trading, though proprietary trading was by no means entirely blameless. Implementation of the Volcker Rule was delayed until July 21, 2015, and several extensions were granted to banks to exit their illiquid investments. By June 2017, the Volcker Rule was under siege as legislators and the president sought to roll back regulations. Alternatives to the Rule at the time included increasing equity capital ratios for banks that engaged in proprietary trading. Nevertheless, many of the larger U.S. banks had already sold or closed proprietary trading desks as early as the end of 2011.

Post Dodd–Frank: The JOBS Act

As lawmakers attempted to hasten improvement of the U.S. economy following the recession, regulatory progress after Dodd–Frank included efforts intended to reduce the substantial costs on raising capital. For example, President Obama signed into law

the Jumpstart Our Business Startups (JOBS) Act in 2012. The JOBS Act was intended to facilitate the so-called emerging-growth companies (capitalizations less than $1 billion) by creating a "mini-registration" process, and allowing for crowdfunding offerings (less than $1 million) with eased reporting requirements. The SEC followed with amendments to Rule 506 of Regulation D and Rule 144A, to permit small investors to participate in such "unregistered offerings" along with the previously allowed accredited investors (investors with income levels exceeding $200,000 or net wealth exceeding $1,000,000 or who hold certain professional securities certifications) and qualified institutional buyers. Such investments by nonaccredited investors would be limited to the smaller of $2000 or 5% of annual income or net worth.

5.7 GOVERNMENT OVERSIGHT OF SELF-REGULATION

The U.S. securities regulatory system might be characterized as a shared, cooperative coordination of industry, state, and federal systems that serve as complementary components under the authority of the U.S. federal government. However, the extent to which these systems share authority and are coordinated might be subject to some debate. Nevertheless, all these systems do play a role in the regulation of securities markets, and the U.S. government can be regarded as the "senior partner."

As discussed previously, 1930s securities legislation provided for some degree of self-regulation in securities markets. While FINRA, exchanges and other markets, and self-regulatory organizations (SROs) themselves still play a primary role in regulating and monitoring market participants, the SEC and other government agencies retain ultimate regulatory powers and frequently play oversight and technical roles in regulatory activity, and seek to "cover gaps" in self-regulation. SROs such as the NYSE have regularly pursued actions against insider trading. If markets fail to properly monitor and regulate themselves, the SEC or CFTC can take action, as might state authorities.

One important example of SRO failure and SEC oversight was the 1990s Nasdaq so-called "odd-eighths" scandal. Nasdaq market makers declined to post customer quotes in odd-numbered eighths, intentionally widening spreads by using round dollar amounts, 1/4, 1/2, and 3/4 fractional amounts (Christie & Schultz, 1994). This, of course, impaired liquidity, increased spreads and brokerage commissions along with dealer trading profits. Over 30 securities firms and individuals settled with the SEC by paying $27 million in fines and incurring other penalties, and agreed to pay an additional $1.03 billion in a federal class action suit for price-fixing. Furthermore, in 1997, Nasdaq agreed to implement a series of revised Order Handling Rules, with three essential components. First, the Limit Order Display Rule requires that sizes and prices of customer limit orders be displayed in the NBBO when they were better than market maker quotes. Second, the Quote Rule requires market makers to publicly display their best quotes, further requiring them to display and honor quotes from ECNs. Third, the Actual Size Rule reduced the minimum displayed quote size from 1000 to 100 shares.

In an earlier example of SRO failure, Richard Whitney, the president of the New York Stock Exchange, was discovered in 1937 to have misappropriated large sums from the NYSE "widows

and orphans" fund to save his own brokerage firm. After this scandal, the SEC required that the NYSE to reorganize itself and strengthen the independence of its board of directors.

Competition among regulatory agencies can enhance the detection of unethical and illegal behavior. In some eras, the SEC has been considered to have been either lax or weak in its enforcement. In certain instances where the SEC has failed to detect market abuses or take corrective action, other regulatory authorities such as state attorneys general have done so. State "blue-sky" laws such as New York State's Martin Law, as we discussed earlier in Section 5.1, have been used by state attorneys general to pursue abuses. In 2016, the U.S. Department of Labor issued regulations requiring brokers paid to provide investment guidance for certain retirement investment accounts to act solely in the best interest of their clients. These regulations have been hotly contested. Other economies employ multiple regulatory agencies as well. For example, securities regulation in the Europe is administered by E.U. and member country government authorities as well as by self-regulatory organizations (SROs) such as securities exchanges.

In addition to contributing to detection efforts, regulatory competition also has the potential to contribute to product and process innovation in financial markets as firms seek their preferred regulatory regimes (Romano, 2001). However, overlapping regulatory regimes can also cause difficulties such as costly duplication in regulatory filings and regulatory "double dipping" on fines. For example, both the SEC and FINRA extracted $2.5 million settlements from Scottrade in 2014/15 over data integrity issues.

The Securities and Exchange Commission

The SEC was created as an independent agency by the Securities and Exchange Act of 1934 to protect investors, to maintain fair, orderly, and efficient markets, and to facilitate capital formation, particularly in the business sectors. The SEC seeks to ensure that firms and organizations raising money by selling securities to investors disclose certain essential facts about these securities prior to their sale and while they are held. The SEC also seeks to ensure that those who trade securities are dealt with fairly and honestly.

The president of the United States, with advice and consent from the Senate, appoints the five members of the SEC, including its chair, to staggered five-year terms. No more than three can belong to any one political party. The Commission is comprised of five divisions and 16 offices, with headquarters in Washington, DC, and in 11 regional offices around the country. The Commission interprets and enforces federal securities laws, and issues and amends rules to complement, implement, and support federal legislation. In addition, the SEC monitors and sanctions securities markets (including firms, brokers, advisors, SROs, auditors, and rating agencies), oversees their inspection, and coordinates activities with state and foreign regulators.

The SEC not only investigates regulatory infractions and enforces legislation, but is an essential rule maker itself. For example, despite legislative and SEC efforts to limit trading on the basis of nonpublic information, many investors still openly did so. For example, it had been common practice for firms to assemble select securities analysts to make corporate announcements to them ahead of announcements to the general public. This type of activity certainly tilted the playing field in the favor of these analysts and their employers.

In response, the SEC introduced *Regulation Fair Disclosure* (FD) in 2000, which was intended to limit this type of selective disclosure. Regulation FD requires companies to make material information public to all investors. Furthermore, if material information were to leak unintentionally, firms are required to reveal the leaked information to all investors within 24 hours.

Enforcement of Securities Regulation

As discussed above, the SEC develops, administers, and enforces securities regulation in the U.S. The Division of Enforcement of the SEC administers the SEC's Enforcement Program, which detects and investigates potential violations of federal securities laws and regulations. Investigations might initiate as either informal, which normally involves voluntary production of documents and interviews, or formal, in which case the Division obtains subpoenas for testimony and document production.

SEC administrative courts are administrative units of the SEC. As such, they do not decide criminal actions, but, as described above, do impose civil and administrative penalties. The Department of Justice pursues both civil and criminal actions through United States Attorneys appointed by the President of the U.S., one for each of 94 federal judicial districts, and their assistant U.S. attorneys. The U.S. Attorney's Office for the Southern District of New York, which includes New York City and nearby counties on the mainland of New York state, prosecutes criminal and civil cases involving violations of federal securities laws in its district. Because of its location, the office has pursued many of the most high-profile securities cases in the U.S., and has a particularly strong reputation for pursuing securities crime.

The SEC and other enforcement agencies obtain information on potential securities violations through a variety of leads, including investor tips, media coverage, market surveillance, analysis of security price histories, referrals from other law enforcement agencies and class action attorneys, etc. The stages of the typical SEC investigation are as follows:

- *Leads*, described in the paragraph above, trigger the investigation.
- *Matter Under Inquiry* (MUI), which is an informal or preliminary investigation to determine whether a formal investigation is warranted and which might consist of as little as a single telephone call and might be closed with no further action taken. MUIs are often accompanied by requests for documents and notices that the individual and/or institution must take steps to preserve all information that might be relevant to the SEC's inquiry.
- *Formal investigation*, during which the SEC enforcement staff will prepare and issue a *Formal Order of Investigation*; review books, records, and trading data; collect witness testimony; review other pertinent information; and seek orders from federal courts enforcing subpoena requirements when appropriate.
- *Wells Process*, which begins once the SEC believes that a violation of the federal securities laws has occurred and determines that it will seek charges. The target of the investigation is delivered a Wells Notice (described below), which is a notification of the SEC's intent to recommend that enforcement proceedings be commenced, potentially followed by the target's Wells Submission responding to the Commission.

- *Enforcement Actions*: After the Wells Process, during which time approximately 20% of investigations are dropped, the SEC may bring a civil case against the target in federal district court. Because the SEC does not have authority to prosecute criminal cases, the SEC refers them to the U.S. Department of Justice or, in fewer cases, to state criminal authorities. The SEC can bring its own civil enforcement actions to federal district court and/or to administrative proceedings.

Early in an investigation process, the SEC is likely to informally approach a target or witness for an informal interrogation. To the extent possible, the Division of Enforcement ascertains facts through interrogations of investigation targets, interviewing witnesses, reviewing telephone and other communications records, reviewing financial records and trading data, etc. When appropriate, the Division of Enforcement (or FINRA) normally delivers to the subject of an investigation a *Wells Notice*, a notification of its intent to recommend, based on its investigation, that enforcement proceedings be commenced. The Wells Notice (named for John Wells, chair of a 1972 committee formed to review the Commission's enforcement) briefly summarizes the violation that the Division of Enforcement is investigating or has investigated. The respondent is allowed to respond to the Wells Notice through a Wells Submission to the Commission. After its investigation, the Division of Enforcement presents its findings to the Commission for review, after which, the Commission can authorize filing an administrative action and/or filing a complaint in a federal district court, or drop the matter altogether.

After investigation by its Division of Enforcement, should the SEC decide to further pursue the matter, it will choose whether to:

1. File an administrative action before an SEC administrative judge presiding over an SEC administrative court. Such an administrative enforcement can lead to enjoining an individual and/or firm from further violations, disgorgement of illegal profits, cease and desist orders, barring or suspending individuals and firms from the securities business, and imposing civil penalties, and/or
2. Recommend civil action and/or criminal prosecution to the Securities and Financial Fraud Unit of the Department of Justice, which prosecutes securities, commodities, and other financial fraud cases, including market manipulation and insider trading. Such actions are pursued through federal district courts.

We discussed earlier in this chapter the roles of the Attorney General for New York State in pursuing violations of New York securities laws, both as civil and criminal procedures. The New York County District Attorney's Office (district attorney for Manhattan) has somewhat analogous powers as the local authority enforcing state financial crime statutes. The Commodity Futures Trading Commission, discussed below, enforces its regulations in a manner similar to that of the SEC.

The Commodity Futures Trading Commission

The CFTC was created as an independent agency in 1974 after the enactment of the Commodity Futures Trading Commission Act to regulate U.S. commodity futures and option markets. Its creation was delayed in comparison to that of the SEC because most

futures and options trading were related to agricultural commodities and thus were overseen by the Department of Agriculture. A number of market developments, including the 1973 opening of the Chicago Options Exchange, led to the 1974 enactment of the Commodities Futures Trading Commission Act.

In a manner similar to the SEC, the CFTC has five commissioners including its chair appointed by the president to staggered five-year terms, with no more than three commissioners from any one political party. Its mission is to protect market participants and the public from fraud, manipulation, and abusive practices related to futures and options, and to foster open, competitive, and financially sound markets. The CFTC maintains a number of essential offices and divisions, including the following[9]:

- The CFTC Office of the General Counsel is the Commission's legal advisor, represents the Commission in appellate litigation and certain trial-level cases, including bankruptcy proceedings involving futures industry professionals, and advises the Commission on the application and interpretation of the Commodity Exchange Act and other administrative statutes.
- The Office of the Executive Director formulates and implements the management and administrative functions of the CFTC and the agency's budget.
- The Division of Clearing and Risk oversees derivatives clearing organizations (DCOs), the clearing of swaps, futures, and options on futures, and market participants that may pose risk to the clearing process.
- The Division of Market Oversight is responsible for fostering markets that accurately reflect the forces of supply and demand for the underlying commodity and are free of abusive trading activity, oversees trade execution facilities, and performs market surveillance, market compliance, and market and product review functions.
- The Division of Enforcement investigates and prosecutes alleged violations of the Commodity Exchange Act and Commission regulations. Violations may involve commodity futures or option trading on U.S. futures exchanges or the improper marketing and sales of commodity futures products to the general public.
- The Office of the Chief Economist provides economic support and advice to the Commission, conducts research on policy issues facing the agency, and provides education and training for Commission staff.

5.8 IMPACT OF REGULATORY ACTIVITY

Successful financial regulation can prevent market failure, promote macroeconomic stability, protect investors, and mitigate the effects of financial failures on the real economy. Financial regulation can also be used to improve market transparency. However, financial regulation also imposes a variety of costs on regulated firms and the economy:

- Direct costs: Costs of regulation administration and enforcement, which might be complicated by having multiple agencies enforce regulation. These costs might be

[9]Most of these CFTC operating unit descriptions were paraphrased from the Commission's website at http://www.cftc.gov/About/CFTCOrganization/index.htm.

financed by some combination of taxes and fees imposed on regulated institutions and their clients as well as taxes imposed on the general public. These costs might also be absorbed by regulated firms through self-regulation.
- Indirect costs: Costs of compliance, such as those associated with maintaining records, hiring compliance officers, making payments to auditors and ratings agencies, and so on.
- Distortions to financial markets: Reactions to regulation often cause institutions to conduct business sub-optimally, and can cause firms to leave or restrict their entry into the marketplace. Regulated firms often seek operating jurisdictions with the least restrictive regulations.

These costs hurt profits of regulated firms, and many are ultimately passed on to the real economy. Even with the financial market crisis of 2007–2009, the U.S. financial regulatory systems are often considered to be among the most highly developed, comprehensive, effective, and mimicked in the world. But these systems are far from perfect, and numerous studies have been conducted to more closely examine the effects of financial regulation.

Firms often respond to regulatory changes (and tax increases) by seeking more attractive jurisdictions in which to operate. This practice is called *regulatory arbitrage*, though it is not really arbitrage under a strict definition. The "race to the bottom" occurs when jurisdictions respond by reducing regulatory (or tax) burdens so as to compete with other jurisdictions to attract business. Hedge funds regularly domicile offshore in order to reduce tax and regulatory burdens. On the other hand, swap dealers and other nonexchange market participants often prefer to trade with counterparties whose jurisdictions provide "pro-creditor" protections. Such regulatory protections can enhance markets.

Macey (2003) proposes *regulatory globalization* as a means through which regulatory agencies can extend their reach. Regulatory globalization involves cooperation between regulatory agencies in different countries (jurisdictions) through agreements or international regulatory institutions such as the G20, IOSCO (discussed in the next section), or the Bank for International Settlements to agree on regulatory regimes.

Sarbanes-Oxley Impact

Since its enactment in 2002, Sarbanes-Oxley (SOX) and its costs and benefits have been a source of controversy (Coates & Srinivasan, 2014). Many observers (see Coates & Srinivasan) argue that the Act did not go far enough in its requirements while many corporate managers argued that the Act imposed unduly expensive reporting and compliance requirements on them. Nonetheless, Cohen, Dey, and Lys (2008) found that SOX did lead to a decline in earnings management, and Gordon, Loeb, Lucyshyn, and Sohail (2006) found that SOX improved the voluntary disclosure of information security activities by firms. Costs of regulation are difficult to measure, and the benefits are usually even more difficult to identify and measure. Based on several extrapolations, Dyck, Morse, and Zingales (2014) estimated the annual benefits of SOX to equal $208 billion net of compliance costs. We will further discuss the benefits of regulation and costs of regulatory violations here and in Chapter 13.

While the benefits and costs associated with SOX continue to be evaluated, U.S. regulatory actions have a number of interesting results, many of which are unintended. First, financial regulatory activity tends to impose significant burdens on business and financial institutions. A Financial Executives International Survey (2004) estimated the mean compliance costs to U.S. business of Section 404 of SOX dealing with internal control quality to be $3 million per firm per year, not counting the distortion costs associated with impacts on company operations. Accounting for the full impact of SOX, including indirect costs, Zhang (2007) estimated through stock market reactions that the total costs imposed on business by SOX compliance would be as high as $1.4 trillion, though, based on survey data, Hochberg, Sapienza, and Vissing-Jorgensen (2009) estimated annual compliance costs to be $6.8 billion per year. While the lower figure is annual, there is still a wide spread between the two cost estimates. Within three years of the enactment of SOX, a number of European companies, including ITV and United Business Media, had deregistered their firms' stock in the United States to avoid these costs. These and other distortions were clearly unintended by the authors of SOX. Vermont Teddy Bear, a Shelbourne, Vermont–based gift and toy company, took its shares private in a highly publicized 2005 act intended to avoid costs of complying with SOX. More generally, Chhaochharia and Grinstein (2007) argue that smaller firms suffer disproportionately from the costs of implementing Sarbanes-Oxley, particularly the costs of maintaining internal controls. In addition, smaller firms experience more difficulty attracting qualified independent directors to their boards. Chhaochharia and Grinstein found that larger firms that needed to make the most significant adjustments to comply with Sarbanes-Oxley experienced the most significant short-term benefits to their share prices.

While SOX has been blamed for causing firms to refuse to list or withdraw from listing their shares in the United States, it is important to note that companies were regularly delisted from the NYSE prior to SOX for a variety of reasons. For example, firms either delist or refuse to list their shares because they simply cannot fulfill or continue to fulfill NYSE listing requirements. Some firms that delisted might simply have found it convenient to blame SOX for being forced into exchange delisting.

Another effect of financial regulation is its encouragement of financial innovation. Numerous financial products and markets owe their creation and/or growth to financial regulation and efforts to bypass regulation. A small sampling of examples includes preferred stock, interest rate swaps, hedge funds, loan securitizations and swap funds. Another innovation from SOX is markets for unregistered shares. Consider GSTRuE, the Goldman Sachs Tradable Unregistered Equity OTC Program, which was, in part, created to sell shares of Oaktree Capital Management, LLC, an alternative management firm. Oaktree, like other alternative management firms Blackstone and Fortress, prefers to operate under regulatory "radars" and with a significant degree of confidentiality. Blackstone and Fortress, as public firms, are subject to SOX and other regulation applicable to public firms. Goldman Sachs floated shares of Oaktree stock to fewer than 499 institutional investors with capital in excess of $100 million, and created GSTRuE so that these institutional investors could trade those issued shares outside of the public securities regulatory reach.

Impact of Reg NMS

Reg NMS (See Section 5.5), in particular, the Order Protection Rule requiring executions at the best quote regardless of quote reliability or size, has been controversial since it was proposed. While Reg NMS did seek to improve market competition and automation, and did seek to protect investor orders, the Order Protection Rule component of Reg NMS has led to serious unintended consequences, as was predicted before its passage. Some observers (e.g., O'Hara, 2004) have noted that Regulation NMS does not account for the diverse needs of different types of traders and has led to a deterioration of liquidity. This deterioration of liquidity has led some traders seeking to bypass better quotes on the NYSE for speedier trades or more anonymous trades in dark markets, further increasing market fragmentation and complexity, thereby increasing market fragmentation. For example, some trades might have executions forced into less reliable or slower markets in order to fulfill the best price requirement while HFT and other traders have learned to game the system brought on by the Order Protection Rule. Redler (2010) and Chung and Chuwonganant (2010) argued that Regulation NMS has, contrary to its objectives, increased spreads and trading costs while reducing liquidity and depth in equity markets. In fact, Redler argued that the May 6, 2010 "flash crash was a direct result of false liquidity in the marketplace insulated by the Order Protection Rule," where HFTs retracted their orders immediately after observing the one-sided market.

Dodd–Frank and the Volcker Rule: Impact on Trading

While much of Dodd–Frank was concerned with banking, stability, and consumer protection, many of its trading components such as the clearing mandate produced somewhat less controversy. A major exception to this was the Volcker Rule, which was largely in effect by mid-2015 as part of Dodd–Frank implementation. In effect, this rule prohibits bank proprietary trading by withdrawing federal assistance (bailout) expectations from banks undertaking risky trading activities, reducing the moral hazard problem (taking risks with the money of others). Hence, many banks spun off or shut down proprietary trading desks. The banking industry expressed much dissatisfaction with the rule because of its negative impact on an important but volatile source of bank profits. In addition, consider the effect of the Volcker Rule on banks' ability to trade on behalf of clients. For example, banks need to maintain and regularly adjust inventories of currency so as to support and maintain markets for clients who trade currencies. Proprietary trading is important to banks to fulfill this key function on behalf of clients.

A 2011 study conducted by Oliver Wyman (a consultancy firm) on behalf of the Securities Industry and Financial Markets Association, a trade group for the securities industry, claimed that the Volcker Rule could cost investors in U.S. corporate bonds as much as $315 billion, due to reductions in liquidity, increases in interest rates, and other distortions caused by the rule. However, many of these losses by banks are likely to be recaptured by other investors and institutions that fill affected banks' proprietary trading roles. In addition, some of these lost profits will be compensated by investors who receive higher interest payments.

The *Economic Growth, Regulatory Relief, and Consumer Protection Act of 2018* was a modest follow-up to Dodd-Frank, intended to provide for regulatory relief, particularly for small banks, and to provide for certain consumer disclosure requirements. The Act redefined the "too-big-to-fail" cut-off for banks at $250 billion (up from $50 billion), reduced the number of systemically important banks from 38 to 12, and subjected only larger banks (>100 billion in assets) to stress testing requirements. It also eliminated the Volcker Rule for banks with less than $10 billion in assets, provided that their trading assets and trading liabilities are less than 5% of total assets.

Implementation of the Volcker Rule proved controversial and remains so. The Fed, FDIC, OCC, SEC, and CFTC finalized regulations for implementation of revised versions of the Volcker Rule in 2020-21. The version of the rule now to be followed by banks is customized, depending on the size of the bank, the bank's level of traded assets, the bank's rationale for trading, and the types of assets that the bank trades.

5.9 REGULATION: THE INTERNATIONAL ARENA

Almost 30 years ago, Benn Steil offered the following transaction description to illustrate the increasing globalization of securities markets:

> An American broker in London buys German "Bunds" for a Mexican client from a Japanese trading house in Paris and clears the trade in Brussels. This sort of story line is not simply plausible, but is on the verge of becoming commonplace (Steil, 1994, p. 197).

However, most securities market regulation (other than certain instances in the European Union) is conducted at the national level, or in some cases, including Canada, even at the provincial level. Regulatory activity on the global level has been more pervasive in the banking industries than in the securities industries.

IOSCO

The most significant nonbanking global organization engaged in regulatory policy is the *International Organization of Securities Commissions* (IOSCO). This organization cooperates in "developing, implementing and promoting adherence to internationally recognized and consistent standards of regulation, oversight and enforcement in order to protect investors, maintain fair, efficient and transparent markets, and seek to address systemic risks."[10] Thus, the organization seeks to improve financial efficiency and transparency, protect investors, and reduce systemic risks. The Organization is comprised of representatives from over 100 national securities regulatory commissions around the world.

Basel

Because of significant overlaps between the banking and securities industries, the *Basel Committee on Banking Supervision* is also a crucial global body providing for regular

[10]Taken from the IOSCO website, http://www.iosco.org/about/.

cooperation on securities regulation. Neither of these organizations has formal regulatory authority in any country, although their policy recommendations are regularly adopted by countries. While the Basel Committee is primarily concerned with bank safety, financial trading does impact bank risk, and Basel directives can affect trading. For example, consider the Basel III directive to vary capital requirements as trading activities impact bank risk. This directive will impact bank trading activities, and banks will incur costs to comply with the requirements.

ESMA

At a more regional level, the *European Securities and Markets Authority* (ESMA) is the European Supervisory Authorities (ESA) body that regulates EU securities markets. Unlike the IOSCO, the ESMA is the only supranational securities regulatory body that has the authority to draft legally binding technical standards and ban securities market activities likely to increase systemic risks, and has the ability to launch fast-track country-specific procedures to ensure consistent application of EU law. The ESMA can initiate investigations and request that EU member countries launch investigations, and can impose fines and issue recommendations based on the results of those investigations. The ESMA also has binding mediation powers to resolve securities market conflicts among member countries.

MiFID

The *Markets in Financial Instruments Directive* (MiFID) was approved in 2004 and took effect in 2007 as the cornerstone of the E.U.'s regulation of financial instruments and markets. It was intended to create a single market based on competing trading venues. In some respects, MiFID is analogous to the National Market System in the U.S. (The Securities Acts Amendments of 1975, discussed in Section 5.4). MiFID provides for standardized rules on the issue of securities, transparency and reporting requirements, prevention of market abuse, client order handling (including best execution), and conduct of securities firms. In order to provide for client protection rights, MiFID categorizes securities firms' clients as follows:

- Retail Clients: Clients not categorized as Professional Clients or Eligible Counterparties
- Professional Clients: Other than Eligible Counterparties—"large undertakings" with 2 or 3 of the following: balance sheet totaling at least EUR 20 million, net turnover of at least EUR 40 million or capital of at least EUR 2 million; or has requested and been granted Professional Client status
- Eligible Counterparties (ECP): Investment firms, credit institutions, insurance companies, other financial institutions, central banks, and national governments

These client groups are granted, in descending order, levels of protection, as professional clients and ECPs are considered to be more experienced and sophisticated. Protections concern the delivery of investment advice, investment suitability assessment, and provision of details on fees and commissions received by the securities firm.

MiFID II and *Markets in Financial Instruments Regulation* (MiFIR) were adopted by the E.U. in 2014 to take effect in 2018. MiFID II seeks to force more trading into regulated trading venues such as exchanges or other eligible platforms. In addition, it enhances rules for algorithmic trading and HFT conduct (e.g., algo testing), provides for non-E.U. firm access to E.U. markets, facilitates small- and medium-sized firms' access to capital, and increases supervisory powers for regulators. MiFIR, seeking to enhance transparency, sets forth improved reporting requirements for pre- and post-trade data to the general public.

Risk Taking, Clawbacks, and Malus

Most financial institutions face a variety of regulations limiting their risk-taking activities and behavior. Clearinghouses, clearing firms, and brokerage firms, as institutions, make efforts to limit their clients' risk taking and carefully manage their own institutional risk-taking. However, financial institution employee incentive programs vary widely, as do government regulations designed to limit risk-taking by employees. Whereas securities firms normally reward their traders handsomely for profitable trading, they have not consistently imposed commensurate penalties for unprofitable trading activity. This asymmetry can lead to excess risk-taking on the part of individual traders and impose significant risk of institutional failure (see Chapter 13 for examples).

Adapting compensation structures can potentially mitigate traders' motivations to engage in excessive risk-taking. Malus and clawbacks are provisions that can be used in trader compensation packages to limit inappropriate risk-taking. A *clawback* is a compensation provision that allows for already vested or paid employee compensation to be reclaimed by an employer based on financial restatements, reputational damage, gross negligence or other malfeasance on the part of the employee. Clawback provisions can be exercised in the event of losses that are realized by the firm after compensation has been awarded. A *malus* provision in a compensation contract reduces or withholds pending compensation in the event of an undesirable performance outcome, such as a loss induced by risk-taking. It is often interpreted as the opposite of a bonus, where compensation is maintained in escrow and can be retroactively reduced when it becomes clear that employee performance did not merit the pay.

Countries differ with respect to the legality of clawback provisions. Most E.U. financial institutions, including those in Germany, Italy, Belgium, and Portugal, are required to implement clawback clauses in their variable remuneration contracts to material risk-takers and certain senior managers. Spain requires employees to provide express written consent for such clauses to be enforceable. U.K. financial institutions are required to subject material risk-takers to clawbacks for at least 7 years after payment or vesting, rising to 10 years for certain senior managers. On the other hand, although Section 954 of the Dodd-Frank Act authorized SEC-mandated and exchange-enforced clawback rules, as of mid-2021, the SEC has not implemented proposed rules for such requirements. Nonetheless, many U.S. financial institutions have implemented such rules anyway, in some cases at the urging of large institutional shareholders such as CalPERS and BlackRock. Unfortunately, most of these clawback rules have been mostly applicable to senior managers and few such rules have resulted in actual clawbacks.

Selected Country Regulatory Authorities

The *Federal Financial Supervisory Authority* (BaFin) is the German financial regulatory authority under the supervision of the *Federal Ministry of Finance*. BaFin is one of the approximately 30 country-specific regulators belonging to and working with ESMA. BaFin supervises banks, banks, financial services institutions, securities exchanges, insurance firms, pension funds, investment funds, and asset management companies. It also enforces standards of professional conduct in financial institutions.

The U.K. maintains two primary financial regulators over financial service firms, known as the *Financial Conduct Authority* (FCA) and the *Prudential Regulation Authority* (PRA). Funded by the institutions that it regulates, the FCA seeks to proactively ensure that consumers are protected in the marketplace and that markets maintain integrity and function well. The FCA is an independent regulator, supervised by the Treasury, which oversees the U.K. financial system. The PRA, an arm of the Bank of England, seeks to proactively ensure the safety and soundness of financial institutions. Both regulators regulate and supervise commercial banks, securities firms, and insurance companies, and, since 2012, have served to replace the failed Financial Services Authority (FSA).

Securities markets are regulated in Japan by its *Securities and Exchange Surveillance Commission* (SESC), established in 1992 as a commission operating under the authority of the Japanese Financial Services Agency. As is the case with the SEC in the United States, the SESC attempts to ensure that investors receive full disclosure with respect to security issues. Unlike in the United States, the SESC imposes eligibility standards on Japanese firms wishing to make public bond issues. The SESC does not have the authority to prosecute violations. The Commission files its findings and recommendations with prosecutors and the Financial Services Agency.

Whereas the U.S. abandoned its emphasis on blue skies regulation in the 1930s, Canada relies on provincial securities regulations. Each of Canada's 10 provinces and three territories has its own securities legislation and its own securities commission. With the exception of Ontario, provinces recognize securities professionals' registration status under any of the other provincial authorities. This mutual recognition is known as the "passport system." In addition, these provincial commissions appoint representatives that comprise the Canadian Securities Administrators (CSA), a national organization that supports and monitors the brokerage industry. The CSA and provincial regulatory authorities have conferred self-regulatory organization (SRO) status on the Investment Industry Regulatory Organization of Canada (IIROC), the Chambre de la Sécurité Financière (CSF—Quebec only), and the Mutual Funds Dealer Association (MFDA). These SROs have the power to regulate the conduct of securities dealers, including mutual fund dealers, all under the ultimate supervision of the CSA. There has been pressure to create national securities legislation and governmental regulatory bodies, but these efforts are still underway.

5.10 PRIVATIZATION OF REGULATION AND EXCHANGE RULES

Privatization of regulation typically refers to nongovernment creation and adoption of common guidelines to govern the behavior of participants in the system (markets). The

CFA Institute (2007) stated that "the overarching purpose of any self-regulatory group is to keep industry interests aligned with the public interest so as to avoid government intervention and the possibility of more-restrictive regulation."

The Maloney Act of 1938 and the formation of the NASD launched a significant innovation leading towards some degree of privatization in the regulation of securities markets. While governmental organizations still play the dominant role in securities regulation, over time there have been wide variations in levels of privatization. NASD (now merged into FINRA), the National Futures Association, and exchanges themselves act or acted as SROs (self-regulatory organizations). However, their activity is overseen and perhaps even micromanaged by the SEC. The OTC (nonpublic) markets rely on trade associations or nonprofit corporations to prepare uniform contracts and standardized agreements among market participants. For example, in 1987 the *International Swaps and Derivatives Association* (ISDA), which sets standards for derivative contracts, prepared the then 14-page ISDA Master Agreement that has regulated the huge swaps markets. This relatively short Agreement acts as a standardized contract used among participants in these huge markets, and reflects a principles-based regulatory regime for the regulated markets. ISDA and its early success were instrumental to the refusal by Congress to further regulate derivatives markets in the 1990s. When such efforts fail to prevent market stress or failure, government regulators intervene. Major advantages to maintaining private regulatory bodies or self-regulation in securities markets might be as follows:

- Market participants have the most intimate knowledge of the markets to be regulated.
- The regulatory foci on developing best practices and effective monitoring and enforcement policies are based on economic and reputational self-interest.
- Governmental regulatory costs are reduced as they are passed on to the regulated market.

NYSE Listing Standards

Exchanges are established to facilitate and regulate trading, and remain important sources of securities markets regulation. Such regulations ensure that all participants remain reliable partners for trading, as well as establish exactly what is to be traded. Exchange rules are intended to ensure that markets remain active and liquid. For example, the NYSE imposes a variety of rules on trading, member firms, and listed firms. While the NYSE considers each application to list company shares for trading on its own merit, its listing standards are considered rather stringent (although standards do vary for NYSE, NYSE American (formerly NYSE MKT), and NYSE ARCA listings). Such standards for NYSE typically include a global market capitalization level ($200 million in many scenarios as of 2017) and at least 1.1 million outstanding shares with at least 400 shareholders holding a minimum of 100 shares, as well as minimum historical earnings levels and restrictions concerning corporate governance (which, if adopted, will include the proposed SEC clawback requirements that we discussed in the previous section). There are a number of other listing requirements in addition to requirements for delisting a security.

NYSE Circuit Breakers and Trading Halts

Circuit breakers are intended to slow or stop market activity so that traders can digest information and react rationally to market fundamentals. For example, with Rule 80A, the NYSE instituted a system of "circuit breakers" following the stock market crash of 1987. This rule followed the Brady Report recommending circuit breakers as a means to limit credit risk, "level the playing field" for individual investors needing more time to obtain and react to new information, reduce the illusion of perfect liquidity in the face of one-sided volume, facilitate price discovery by giving traders additional time for analysis, inhibit panic, and publicize order imbalances. This rule required that sell (buy) orders related to *index arbitrage* in component stocks of the S&P 500 Index could be executed only on an uptick (downtick) or zero tick.[11] A 2% movement in the DJIA in a single day triggered the rule. The rule was rescinded in 2007.

NYSE Rule 7.12 (formerly Rule 80B) provides for more drastic reactions to more dramatic market volatility. Market-wide circuit breaker triggers, mostly unchanged since 2013, are now as follows:

Level 1: A 7% drop in the S&P 500 Index from its previous close: Trading halts on all stocks on all exchanges for 15 minutes unless it occurs after 3:25 PM.
Level 2: A 13% drop in the S&P 500 Index from its previous close: Trading halts on all stocks on all exchanges for 15 minutes unless it occurs after 3:25 PM.
Level 3: A 20% drop in the S&P 500 Index from its previous close: Trading halts on all stocks on all exchanges for the remainder of the day.

Only one Level 1 and one Level 2 halt can occur in a single trading day.

The highly volatile period caused by selloffs at the start of the 2020 coronavirus pandemic led to 4 market-wide circuit breaker Level 1 closures in March 2020. Afterwards, the "MWCB Task Force," comprised of members from stock exchanges, the SEC, the CFTC, the CME, and several securities industry members, concluded that the circuit breakers generally functioned and served their roles as intended and recommended no major changes.

Earlier in 2010, in response to the May 6 "flash crash," the SEC mandated a uniform single-stock circuit breaker for most of the largest equities irrespective of marketplace, where stocks whose price moves up or down by more than 5% in a five-minute period are subject to a five-minute (or longer) pause in trading activity. These single-stock circuit breakers for individual National Market System (NMS) securities have evolved into a more elaborate set of "moving" price bands known as the *Limit Up-Limit Down* (LULD) mechanism, which restricts NMS security prices to remain within trading bands. These bands are set by the appropriate SIP (SIAC or UTP) based on the security's average price over the past 5 minutes and updated by the SIP every 30 seconds. The widths of the bands depend on the stock's price activity and time of day, normally between 5% and 20%. The bands' widths are based on the average security price over the past 5 minutes, or the daily open early in the day or reopen price after a halt. All executable orders and quotes must fall within the allowable trading bands set for the security. When there are no executable

[11]Index arbitrage involves simultaneous transactions of a large number of securities comprising a particular index, in this case, the S&P500.

quotes for a given security from any exchange over a 15-second period, the market for that security halts for 5 minutes and the bands are reset based on a reopening auction price obtained by the listing exchange.

Consider a somewhat simplified numerical example illustrating the Limit Up-Limit Down mechanism. Suppose that 10% price bands have been set for a given NYSE stock with an average transaction price of $100 over the past five minutes. For the next 30 seconds, only quotes between $90 and $110 are executable for the stock. Now, suppose negative information about the company's stock is released. If the stock should fail to receive any quote within this 10% range from any trading venue for 15 seconds, all trading venues must pause trading this stock for 5 minutes. Suppose that the reopening auction conducted at the end of the halt by the NYSE produced a price of $80 for the stock. Then the trading bands will be re-set for 30 seconds between $72 and $88.

Despite the MWCB Task Force conclusions described above, the extent to which these circuit breakers have actually benefited the market is not clear. First, traders complain that if market price begins to approach circuit breaker levels, their changes tend to accelerate towards the level needed to trigger the circuit breaker (magnet effect). Second, the extent to which the exchanges coordinate their circuit breaker activity is not clear.

Additional Reading

Kitch (2000) and Niemeyer (2001) are very readable and reasonably short introductions to the regulation of securities markets along with considerations for improving regulation. Congressional Research Service (2020) provides an easy to read and slightly more general introduction to financial regulation. Fleming (2011) and the first chapter of Loss and Seligman (2004) present interesting early histories of securities regulation. Websites for the NYSE, FINRA, CFTC, and SEC all provide substantial details on regulations, including the original documentation. The same is true for most of the international and private securities regulatory bodies. The Securities and Exchange Commission Historical Society (2016) has posted a very readable timeline on the history of financial regulation at http://www.sechistorical.org/museum/timeline/1930-timeline.php. The FINRA site for the Series 7 Examination is a good starting point for readers interested in pursuing careers in the securities industry (http://www.finra.org/Industry/Compliance/Registration/QualificationsExams/RegisteredReps/P011051). SecLaw.com provides an extensive collection of articles on securities regulation, ranging from the very introductory to the minutiae of specific securities and markets. Ford (2008) offers a nice discussion of principles-based and rules-based securities regulation.

References

CFA Institute. (2007). *Self-Regulation in today's securities markets: Outdated system or work in progress?* (Vol. 2007, No. 7). Codes, Standards and Position Papers. CFA Institute Centre For Financial Market Integrity, September.

Chhaochharia, V., & Grinstein, Y. (2007). *Corporate governance and firm value: The impact of the 2002 governance rules*. Johnson School Research Paper Series No. 23-06, AFA 2006 Boston Meetings Paper, March.

Christie, W. G., & Schultz, P. H. (1994). Why do NASDAQ market makers avoid odd-eighth quotes?. *Journal of Finance, 49*(5), 1813–1840.

Chung, K.H., & Chuwonganant, C. (2010). Regulation NMS and market quality. *Financial Management*. Available at: http://dx.doi.org/10.1111/j.1755-053X.2012.01184.x

Coates, J. C., IV, & Srinivasan, S. (2014). SOX after ten years. *Accounting Horizons, 28*(3), 627–671.

Cohen, D., Dey, A., & Lys, T. (2008). Real and accrual-based earnings management in the pre- and post-Sarbanes Oxley periods. *Accounting Review, 82*(3), 757–787.

Congressional Research Service (2020). Who regulates whom? An overview of the U.S. financial regulatory framework. Unpublished CRS report. Available at: <https://crsreports.congress.gov/product/pdf/R/R44918> Accessed 09.09.21.

Cravath, Swaine and Moore. (2001). *Memorandum for ISDA members: Commodity futures modernization act of 2000*, New York, NY, January 5, 2001. Available at: <http://www.google.com/url?sa=t&rct=j&q=2(h)%20and%20(g)%20of%20the%20commodity%20futures%20modernization%20act%20of%202000&source=web&cd=3&sqi=2&ved=0CDMQFjAC&url=http%3A%2F%2Fwww.isda.org%2Fspeeches%2Fpdf%2FAnalysis_of_Commodity-Exchange-Act-Legislation.pdf&ei=qiZgT9-tNIrz0gG18piyBw&usg=AFQjCNFdfROr92VJZRfKcF_G2Evyw7TKvQ> Accessed 13.03.12.

Dyck, A., Morse, A., Zingales, A. (2014). *How pervasive is corporate fraud?* Working paper. Available at: <http://faculty.haas.berkeley.edu/morse/research/papers/DyckMorseZingalesPervasive.pdf> Accessed 20.09.21.

Financial Executives International (FEI). (2004). *Section 404 costs survey*. <https://www.sec.gov/info/smallbus/acspc/fei.pdf> Accessed 13.06.17.

Fleming, R. A. (2011). 100 years of securities law: Examining a foundation laid in the Kansas Blue Sky. *Washburn Law Journal*, *50*(3), 583–610.

Ford, C. L. (2008). New governance, compliance, and principles-based securities regulation. *American Business Law Journal*, *45*(1), 1–60.

Gordon, L. A., Loeb, M. P., Lucyshyn, W., & Sohail, T. (2006). The impact of the Sarbanes-Oxley Act on the corporate disclosures of information security activities. *Journal of Accounting and Public Policy*, *25*, 503–530.

Hed-Hofmann, T. (1965). The Maloney Act experiment. *Boston College Industrial and Commercial Law Review*, *6*(2), 187–218.

Hochberg, Y. L., Sapienza, P., & Vissing-Jorgensen, A. (2009). A lobbying approach to evaluating the Sarbanes Oxley Act. *Journal of Accounting Research*, *47*(2), 519–583.

Kitch, E. (2000). Regulation of securities markets. In B. Bouckaert, & G. De Geest (Eds.), *Encyclopedia of law and economics. The regulation of contracts* (Vol. III). Cheltenham: Edward Elgar. Available at: <http://encyclo.findlaw.com/5660book.pdf> Accessed 05.01.12.

Langevoort, D. C. (2020). Watching insider trading law wobble: Obus, Newman, Salman, two Martomas, and a Blaszcak. *Fordham Law Review*, *89*, 507–528.

Loss, L., & Cowett, E. M. (1958). *Blue sky law*. New York: Little, Brown and Company.

Loss, L., & Seligman, J. (2004). *Fundamentals of securities regulation* (5th ed.). New York: Aspen Publishers.

Macey, J. R. (2003). Regulatory globalization as a response to regulatory competition. *Emory Law Journal*, *52*, 1353–1379.

Macey, J. R. (2017). Beyond the personal benefit test: The economics of tipping by insiders. *Journal of Law and Public Affairs*, *2*(2).

Mahoney, P. G. (2015). *Wasting a crash: Why securities regulation fails*. Chicago: University of Chicago Press.

Markham, J. W. (2010). The subprime crisis—A test match for the bankers: Glass-Steagall vs. Gramm-Leach-Bliley. *University of Pennsylvania Journal of Business Law*, *12*(4), 1081–1134.

Niemeyer, J. (2001). *An economic analysis of securities market regulation and supervision: where to go after the Lamfalussy report?* SSE/EFI Working Paper Series in Economics and Finance No. 482. Stockholm, Sweden: Economic Research Institute, Stockholm School of Economics.

O'Hara, M. (2004). Searching for a new center: U.S. securities markets in transition. *Federal Reserve Bank of Atlanta Economic Review*, 37–52.

Oliver Wyman, Inc. (2011). *The Volcker rule: Considerations for implementation of proprietary trading regulations*. Available at: Unpublished report, <http://www.sifma.org/workarea/downloadasset.aspx?id=22888> Accessed 04.01.12.

Redler, S. (2010). *SEC should rethink reg NMS to fix HFT liquidity problem*. MrSwing. Available at: <http://www.mrswing.com/articles/SEC_Should_Rethink_Reg_NMS_to_Fix_HFT_Liquidity_Pr.html> Accessed 04.01.12.

Romano, R. (2001). The need for competition in international securities regulation. *Theoretical Inquiries in Law*, *2*.

Schlesinger, A. M., Jr. (1958). *The coming of the New Deal: 1933–1935, The Age of Roosevelt* (Vol. II). New York: Houghton-Mifflin.

Securities and Exchange Commission Historical Society. (2016). *Timeline*. <http://www.sehistorical.org/museum/timeline/1930-timeline.php> Accessed 06.12.17.

Sgambati, B. (2016). Using the Martin Act to bring fraudulent practices in dark pool promotion to light: An analysis of the Martin Act's applicability to misrepresentations regarding the operation of dark pools. *Columbia Journal of Law & Social Problems*, *Summer*, 585–610.

Steil, B. (1994). International securities markets regulation. In B. Steil (Ed.), *International financial market regulation* (pp. 197–232). West Sussex: John Wiley & Sons.

Strong v. Repide. (1909). *Justia US Supreme Court Center—213 U.S. 419 (1909)*. Supreme.justia.com <http://supreme.justia.com/us/213/419/> Accessed 20.08.12.
Volcker, P. (2012). *Commentary on the restrictions on proprietary trading by insured depositary institutions* <http://online.wsj.com/public/resources/documents/Volcker_Rule_Essay_2-13-12.pdf> Accessed 06.09.17.
Whitehead, C. K. (2011). The Volcker rule and evolving financial markets. *Harvard Business Law Review*, *1*, 39–73.
Zhang, I. X. (2007). Economic consequences of the Sarbanes Oxley Act of 2002. *Journal of Accounting and Economics*, *44*(1–2), 74–115.

5.11 EXERCISES

1. **a.** Some lawyers and economists have argued that legalizing insider trading would benefit markets by allowing insiders to disseminate information through their own trading activity. Prepare an argument to support this position.
   b. Some observers argue that the SEC's efforts to "level the playing field" by limiting insider trading could, in the extreme, ultimately undermine incentives to perform important securities research needed to support market efficiency, further impeding society's efficient allocation of resources. Prepare an argument to support this position.
2. The following was taken from the Securities Exchange Act of 1934, Section 10, amended as of January 3, 2012:

   > It shall be unlawful for any person, directly or indirectly, by the use of any means or instrumentality of interstate commerce or of the mails, or of any facility of any national securities exchange. . .. To use or employ, in connection with the purchase or sale of any security registered on a national securities exchange or any security not so registered, or any securities-based swap agreement, any manipulative or deceptive device or contrivance in contravention of such rules and regulations as the Commission may prescribe as necessary or appropriate in the public interest or for the protection of investors.

   This rule has been used as a basis to prosecute insider trading activity. Describe how regulator reliance on this rule might characterize an example of "principles-based" regulation in the United States.
3. **a.** Insider trading law has been described as a field largely belonging in the arena of common law rather than arising from statutes (e.g., Langevoort 2020). Why might insider trading in some respects be said to belong in the arena of common law?
   b. Describe how the SEC and the Supreme Court of the United States approaches to the legality (or illegality) of insider trading differ.
4. In response to the Financial Crisis of 2007-09, Congress passed major financial system regulatory reforms known as the Dodd–Frank Wall Street Reform and Consumer Protection Act of 2010. What were the primary intentions and provisions of this important legislation?
5. How is the Volcker Rule related to the moral hazard problem in banking institutions?
6. This chapter discussed several of the advantages of maintaining private regulatory bodies for securities markets. What might be some of the disadvantages to this type of securities market regulation?

7. Some people commented that the Securities Act of 1933 was based on the "sunlight theory of regulation," which is analogous to others saying that "those who are forced to undress in public will presumably pay some attention to their figures." What do these commentators mean by such expressions?
8. The United States has maintained a "rules-based" approach to securities regulation while the UK has been progressing to a "principles-based" system of securities regulation.
   a. Consider the rules-based and principles-based regulatory approaches to traffic speed limits. How might speed limit regulations be phrased in either case?
   b. Why might rules-based speed limit regulations be more difficult to implement?
   c. Why might principles-based speed limit regulations be more difficult to enforce?
9. a. Discuss the differences between "best price" and "best execution" with respect to order handling.
   b. Describe how following the SEC's Order Protection Rule (best price) can conflict with "best execution," particularly in the scenario involving a block order.
   c. The "Best Price" component of Reg NMS has been criticized as being counterproductive in that it can serve to prevent brokers from working orders to obtain the best executions for their clients. In Section 1.6, we discussed various types of orders that can be placed in markets. Which of those order types can serve the purpose of obtaining "best execution" when other order types would be executed under "best price" rules?
10. As a shareholder-owned business, the primary responsibility of the NYSE is to its shareholders. As an SRO, the primary responsibility of the NYSE is to maintain the integrity of markets and to protect investors. There are obvious possible conflicts of interest here. However, make an argument that this apparently conflicting set of responsibilities need not impair the ability of the NYSE to fulfill its responsibilities.
11. A non-U.S. company incurs a significant expense due to regulatory and exchange requirements when it lists its stock (or an ADR for its stock) on the NYSE. Furthermore, many such firms will remain regulated by their host countries. Some observers suggest that this significant regulatory expense is necessarily incurred by the listing firm in order to access the deeper U.S. capital markets. What other reasons might a non-U.S. firm have for incurring the significant expenses of listing on the NYSE, perhaps in addition to a market in its host country?
12. What is the primary difference between compensation clawback and malus provisions?
13. The United States Congress and the SEC worked towards and issued Regulation NMS anticipating that traders would post their best quotations, to the benefit of other traders and the marketplace. Have trading participants in the marketplace acted so as to circumvent the good intentions of Congress and the SEC? If so, how?

CHAPTER

6

Adverse Selection, Trading, and Liquidity

6.1 MARKET MICROSTRUCTURE, LIQUIDITY, AND SPREADS

In Section 1.5, we characterized market microstructure as being concerned with the markets for transaction services. In Section 1.6, we noted that liquidity refers to the ability to easily transact without causing significant price changes. However, it is important to appreciate that while liquidity is easily recognized by every trader, it is a somewhat elusive concept, unobservable and cannot be properly measured with a single metric or proxy. Volumes or spreads might be used as simple measures to capture certain aspects of liquidity, but we know from price impact that both are far from being a complete measure of liquidity.

More generally, this chapter is concerned with information, uncertainty, *adverse selection* (a party to a contract using her superior information to the disadvantage of her counterparty) and dealer inventories on trading and market microstructure. We will examine how market microstructure impacts trader behavior and prices. Later in this section, we will examine a simple model depicting how security spreads affect prices and returns. Material in this chapter is of practical importance to traders because it forms an important basis for many algo and HFT trading models in which traders anticipate order flow based on characteristics of existing and past order flow.

Traders are dependent on information, but differences in trader access to information can have perverse effects on liquidity and how securities are priced. We will discuss in Sections 6.2 to 6.6 how traders with inferior information alter their trading behavior in the presence of counterparties with superior information.

Liquidity and liquidity risk play important roles in the pricing of securities (see, for example, Amihud, Mendelson, & Pederson, 2005). Empirical studies verify that reductions in liquidity reduce security prices and increase required returns (e.g., Chan, Hong, & Subrahmanyam, 2008; Chaplinsky & Ramchand, 2004). Brennan and Subrahmanyam (1996) question this positive relationship between asset returns and liquidity, but Brennan, Chordia, Subrahmanyam, and Tong (2012) find that sell-side illiquidity and its effects far exceed buy-side illiquidity and its effects, and that this sell-side illiquidity is reflected in security returns.

Financial Trading and Investing
DOI: https://doi.org/10.1016/B978-0-323-90955-6.00006-9

© 2023 Elsevier Inc. All rights reserved.

In Chapter 4, we discussed liquidity as a function of direct transactions costs and the price impact of trading. First, consider how transactions costs affect the price of a security. We will extend this discussion in this chapter by examining the roles of uncertainty, information asymmetry, and dealer inventories on liquidity. First, we will discuss the effects of liquidity on prices and returns as reflected in spreads.

Asset Pricing and the Bid-Ask Spread

In the absence of uncertainty and dividend growth, the familiar Gordon stock pricing model is:

$$P = \frac{d}{r}$$

where P is the stock value, d is the constant annual dividend payment, and r is the appropriate discount rate. Now, consider a market with transactions costs reflected by a constant relative spread s ($s=(P_a - P_b)/P$), and where the stock is traded η times per year. An investor buys the stock from a market maker at price P_a, sells it to the market maker at P_b, and adjusts his valuation P of the stock accordingly. We revise the Gordon model as follows (See Amihud & Mendelson, 1986):

$$P = \frac{d}{r + \eta s}$$

Essentially, the total annual transactions cost as reflected by the proportional spread is added to the discount rate for purposes of valuation; the value of an asset is reduced by the present value of all of its expected future trading costs. The proportional spread incurred η times per year must be recovered before an investor can break even. For example, a stock traded twice per year with an average spread of 3% paying annual dividends equal to \$1 discounted at 5% is valued under each of our two frameworks as:

$$P = \frac{\$1}{0.05} = \$20 \quad P = \frac{\$1}{0.05 + 2 \times 0.03} = \$9.09$$

Clearly, the spread has a substantial impact on the value of the security. Notice that a reduction in the spread from 3% to 1% would significantly increase the stock price:

$$P = \frac{\$1}{0.05 + 2 \times 0.01} = \$14.29$$

Now, suppose that the discount rate in the absence of a spread for the stock, $r=E[R]=0.05$, is determined by the Capital Asset Pricing Model or other valuation model when there is no liquidity premium. In economies where stock expected returns include liquidity risk premiums, stock expected returns would be $E[R]=r+\eta s$. For our two liquidity risk premium scenarios, expected returns are $E[R]=0.05+2 \times 0.03=0.11$ and $[R]=0.05 +2 \times 0.01=0.07$. Obviously, the absence of liquidity that causes spreads to increase will increase expected (required) returns and decrease prices. In the next several sections, we will discuss factors that affect this spread, including uncertainty, adverse selection, noise trading, dealer inventories, and volatility.

6.2 INFORMATION AND TRADING

The *economics of information* is concerned with how information along with the quality and value of this information affect an economy and economic decisions. Information can be inexpensively created, can be reliable, and, when reliable, is valuable. Some economists have suggested that more than half of the U.S. economy is currently engaged in activities that are producing and analyzing information products. In fact, revenues obtained by the New York Stock Exchange from selling trading data exceed its trading-fee based revenues. The simplest microeconomics models assume that information is costless and all agents have equal access to relevant information. But, such assumptions do not hold in reality, and costly and asymmetric access to information very much affects how traders interact with each other. Investors and traders look to the trading behavior of other investors and traders for information, which affects the trading behavior of informed investors who seek to limit the information that they reveal.

In Chapter 4, we discussed slippage and market impact models, describing how prices move against traders who execute large orders due to the perceived information content of those orders. Here, particularly in our discussion of the Bagehot (1971), Kyle (1985), and Glosten and Milgrom (1985) models, we discuss the market mechanisms causing prices to react to the information content of trades (market impact or slippage), and how traders and dealers can react to this information content to maximize their own profits (or minimize losses). This chapter is concerned primarily with problems that arise when traders and other market participants have inadequate, different (*asymmetric information availability*), and costly access to information. In this chapter, we discuss information and information asymmetries in a trading context.

Adverse Selection

Adverse selection originally referred to the tendency of higher risk individuals to seek insurance coverage. More generally, adverse selection refers to precontractual opportunism where one contracting party uses his private information to the counterparty's disadvantage. For example, the adverse selection problem can arise when a woman planning a pregnancy purchases health insurance, when a car rental customer secretly planning a trip through a Golan Heights minefield buys comprehensive insurance on the car, or when a pyromaniac purchases fire insurance. In all three cases, the agent (insured or customer) has private information with respect to the higher anticipated costs of the insurance coverage or lease, but pays a "pooling" premium for the incident or casualty coverage. Obviously, this private information affects the behavior of insurers and other insured clients, in what might otherwise be taken to be a suboptimal manner, referred to as the adverse selection problem. In a financial trading context, adverse selection occurs when one trader with secret or special information uses that information to her advantage at the expense of her counterparty in trade. Trade counterparties realize that they might fall victim to adverse selection, so they carefully monitor trading activity in an effort to discern which trades are likely to reflect special information. For example, large or numerous buy (sell) orders originating from the same trader are likely to be perceived as being motivated

by special information. Trade counterparties are likely to react by adjusting their offers (bids) upwards (downwards), resulting in slippage as we discussed in Chapter 4. This market impact or slippage is the market's reaction to the adverse selection problem. In this chapter, we will consider the impact of the adverse selection problem on order sizes, security prices, and the spread.

6.3 NOISE TRADERS

N*oise traders* trade on the basis of what they falsely believe to be special information or misinterpret useful information concerning the future price or payoffs of a risky asset. Noise traders make investment and trading decisions based on incorrect perceptions or analyses of the market, perhaps creating opportunities for more sophisticated investors and traders. In many models of markets, noise traders are assumed to have no useful information or at least react inappropriately to useful information concerning a security's fundamentals. But, do noise traders distort prices? If noise traders trade in large numbers, if their trading behavior is correlated, or if their effects cannot be mitigated by informed and rational traders, they may well distort prices from fundamental values. In many models, individual investors are presumed to be noise traders and that their psychology (see Chapter 11) and poor access to information inhibit their trading success.

In 1953, the economist Milton Friedman suggested that traders who produce positive profits do so by trading against less rational or poorly informed investors who tend to move prices away from their fundamental or correct values. Fama (1965) argued that when irrational trading does occur, security prices will not be significantly affected because more sophisticated traders will react quickly to exploit and eliminate any deviations from fundamental economic values. Figlewski (1979) suggested that it might take irrational investors a very long time to lose all their money and for prices to reflect security intrinsic values, but nonetheless, those traders who choose their portfolios irrationally are doomed in the long run.

Noise traders are useful, perhaps even necessary, for markets to function. Without noise traders, markets would be informationally efficient. Prices would always fully reflect information (to the extent that market frictions and liquidity trading are absent). Even with asymmetric information access, informed traders can fully reveal their superior information through their trading activities, and prices would reflect this information and ultimately eliminate the motivation for information-based trading. That is, as Black (1986) argued, without noise traders, dealers would widen their spreads to avoid losing profits to informed traders such that no trades would ever be executed. However, noise traders do impose on other traders the risk that prices might move unpredictably, irrationally, and without reference to relevant information. In some instances, this risk imposed by noise traders might discourage arbitrageurs from acting to exploit price deviations from fundamental values. Prices can deviate significantly from rational valuations, and can remain different for long periods. In fact, arbitrageurs might sometimes need to ask themselves the following important question: "Does my ability to remain solvent exceed the asset price's ability to remain irrational?" We will discuss this issue more in Chapter 7.

6.4 ADVERSE SELECTION IN DEALER MARKETS

Bagehot (1971) described a market where dealers or market makers stand by to provide liquidity to every trader who wishes to trade, losing on trades with informed traders but recovering these losses by trading with uninformed, noise, or liquidity-motivated traders. The market maker sets prices and trades to ensure this outcome, on average. The market maker merely recovers his operating costs along with a "normal return." In this framework, trading is a zero-sum or neutral game. Informed investors armed with special information or superior analytical skills will earn abnormal returns; uninformed investors will lose more than they make. In fact, the more frequently an uninformed trader trades, the more he can be expected to lose, and these trading losses are reflected as informed trader profits.[1] Market makers observe buying and selling pressure on prices and set prices accordingly, often making surprisingly little use of fundamental analysis when making their pricing decisions. Kyle's (1985) theoretical model describes the trading behavior of informed traders and uninformed market makers who set prices in an environment with noise traders.

Kyle: Informed Traders, Market Makers, and Noise Traders

Suppose that two rational traders have access to the same information and are otherwise identical. They have no motivation to trade. Now, suppose they have different information. Will they trade? Not if one trader believes that the second will trade only if the second has information that will enable him to profit at the first trader's expense. Traders will not trade if they expect to lose money to their counterparty in trade. In fact, rational traders might not trade against other rational traders even if their information differs. This is a variation of the Akerlof "lemon problem," in which traders lacking superior information decline to trade with traders who have superior information. So, why do we observe so much trading in the marketplace? Most of us believe that others are not as informed or rational as we are or that others do not have the same ability to access and process information that we do. Or, maybe one of the parties makes price concessions sufficient to cover any informational advantage (e.g., Rock, 1986). In Kyle's analysis, the market maker sets prices so that his losses to informed traders are offset by profits realized from noise trader transactions. Therefore, we are happy to trade with our less informed and capable counterparties.

The Setting: Kyle's Adverse Selection Model

Kyle models edict trading and price setting in a market where some traders are informed and others (noise or liquidity traders) are not. Dealers or market makers, who have information only about net demand for or supply of the stock, serve as intermediaries between informed and uninformed traders, and seek to set security prices that enable them to survive even without the special information enjoyed by informed traders. Kyle models how informed traders will use their information to maximize their trading profits given that their trades yield useful information to market makers. Furthermore, market

[1]See Barber and Odean (2000, 2001, 2002) and Barber, Lee, Liu, and Odean (2009) who find that more aggressive (frequent) traders underperform risk-adjusted benchmarks. We discuss these and related findings in Chapter 11.

makers will seek to learn from the informed trader's trading behavior, and the informed trader will seek to disguise his special information from the dealer by reducing the volume of his trading activity. Successful deception by the informed trader will reduce the price slippage imposed by the market maker who sets the stock price.

Traders rely on the market and on each other for information. Suppose, for example, that the unconditional value of a stock in a Kyle (1985) framework is normally distributed with an expected value equal to $E[v]$ and a variance equal to Σ_0. However, suppose that the informed trader has private information that the value of the stock is actually v and is able to reduce his own uncertainty to zero. Uninformed investor trading is random, normally distributed with an ex-ante expected net share demand of zero ($E[u]=0$; negative net demand represents net selling activity or net supply) and a variance equal to σ_u^2. In the Kyle model, the informed trader and noise traders take the first step and submit market orders to the dealer, who can observe the total level of order volume $X=x+u$, where u reflects aggregate noise trader transactions and x reflects informed demand. However, the dealer cannot distinguish between x and u. The dealer's problem is to set a market clearing price p for the security so that his losses on transactions with informed traders are offset by his profits from trading with noise traders.

Informed Trader Demand

In the context of the Kyle model, what allows the informed trader to maximize trading profits and gives the informed trader the ability to camouflage his trades by having them appear to blend in with noise traders? First, any time that $v > E[v]$ (or, $v < E[v]$), the informed trader's private information indicates a buying opportunity (selling opportunity), and the purchase (sale) of additional shares increases his profits. Thus, in the absence of the market maker inferring information from the informed trader's demand and adjusting market prices accordingly, the informed trader's profits are maximized when x equals ∞ given favorable private information or $-\infty$ given negative information. However, the market maker does observe and use the level of total demand $X=[x+u]$ for the stock. While uninformed demand u is purely random, X is directly, though imperfectly related to x. When the market maker observes a large total demand X, she infers, with error, a high informed demand level x, and accordingly increases the market price p of the stock, decreasing the informed trader's profits. Thus, the larger the informed trader's demand, the more shares he purchases and the greater will be the informed trader's trading profits in the absence of price changes. However, high total demand X for the stock will cause the market maker to increase the stock price p, thereby decreasing the informed trader's profits. In the presence of these offsetting pressures on profits, how does the informed trader set his order size? As we demonstrate in Appendix A to this chapter, the informed trader's demand is linearly related to the informed trader's informational advantage $|v - E[v]|$ but inversely related to the uncertainty Σ_0 resolved by informed trader demand relative to the noise σ_u^2 associated with uninformed trader demand:

$$x = \frac{v - E[v]}{\sqrt{\Sigma_0/\sigma_u^2}} \tag{6.1}$$

Thus, the informed trader's purchase demand for shares increases at a constant rate as his informational advantage $v - E[v]$ and uninformed trader camouflage σ_u increase, but decreases with his anticipation that the market maker expects his uncertainty advantage improves (Σ_0). Similarly, the informed trader's selling supply for shares increases as the absolute value of his informational advantage $v - E[v]$ and uninformed trader camouflage σ_u increase, but decreases with his anticipation that the market maker expects his uncertainty advantage improves (Σ_0).

Market Maker Price-Setting

The informed trader would prefer to purchase or sell an infinite number of shares to exploit his informational advantage, but cannot because the market maker would correctly infer that his transactions convey meaningful information and resolves uncertainty at a rate related to Σ_0/σ_u^2. Thus, the market maker institutes price revisions leading to slippage in the market price p. At what level does the dealer set the market price, given her observation of the total demand X=[x+u]? As we demonstrate in Appendix A to this chapter, the market maker sets a price that is linearly related to total demand X=[x+u] for the stock, directly related to the uncertainty Σ_0 resolved by informed trader demand relative to the noise σ_u^2 associated with uninformed trader demand:

$$p = E[v] + \left(\frac{1}{2}\sqrt{\Sigma_0/\sigma_u^2}\right)(x+u) = E[v] + \frac{1}{2}\lambda(x+u) \tag{6.2}$$

Thus, the market price of the stock set by the market maker equals its expected value plus its total demand multiplied by a sensitivity factor λ related to the proportion of uncertainty presumed to have been resolved by the informed trader. Thus, λ (sometimes referred to as *Kyle's lambda*) is the sensitivity of the dealer price to order flow; $1/\lambda$ measures the depth of the market (the order flow needed to cause the price to change by \$1) for the stock. If the market maker believes that the informed trader uncertainty resolution is large compared to uninformed trader volatility, she will adjust the market price of the stock by larger amounts given total demand x+u.

There is an abundance of empirical evidence on the relationship between order flow and market impact (See for example, Zarinelli, Treccani, Farmer, & Lillo, 2015). While Kyle, due to simplifying assumptions, predicts a positive and linear relationship (positive and constant lambda), statistical evidence seems more consistent with a positive and concave relationship. This concavity might be partly explained by a combination of factors, including the tendency of the informed trader to slice trade sizes and trade over time and dealer risk aversion.

Illustration: Kyle's Adverse Selection Model

Suppose in this example, that the unconditional value of a stock in a Kyle (1985) framework is normally distributed with an expected value equal to $E[v]$=\$50 and a variance equal to Σ_0=30. However, the informed trader has private information that the value of the stock is actually v=\$45 per share. Uninformed investor trading is random and normally distributed with an ex-ante expected net share demand of zero and a variance σ_u^2 equal to

5000. The dealer can observe the total level of order volume $X=x+u$ where u reflects noise trader transactions and x reflects informed demand, but the dealer cannot distinguish between x and u.

Under the market and trading circumstances set forth above, what would be the level of informed trader demand for the stock? Since $v < E[v]$, the informed trader will wish to sell shares, and we solve for x as follows, using Equation (6.1):

$$x = \frac{v - E[v]}{\sqrt{\Sigma_0/\sigma_u^2}} = \frac{45 - 50}{\sqrt{\frac{30}{5000}}} = \frac{-5}{0.07746} = -64.5497$$

Thus, the informed trader will provide a net supply of 64.5497 shares to be sold in the marketplace. Notice that an increase in either the variance of uninformed trader demand σ_u^2 or the expected value of the asset without special information $E[v]$ would increase the number of shares that the dealer informed trader would supply. On the other hand, an increase in either the amount of information that the informed trader's special information resolves Σ_0 or the true value of the asset would reduce the supply by the informed trader.

The informed trader would wish to sell an infinite number of shares to earn a $5 profit on each, but cannot because the market maker would correctly infer that his share sales convey meaningful information, and the market maker's price revisions would lead to slippage. Thus, at what level does the market maker set her price, given the total demand $X=[x+u]=-64.5497+0$ that she observes? Since the order volume X is negative, the market maker will set a price less than its unconditional expected value equal to 50, and we solve for the price level using Equation (6.2) as follows:

$$p = 50 + \frac{1}{2}\sqrt{30/5000}(-64.5497 + 0) = 47.50$$

In this scenario, net uninformed order level was zero. For a given increase in the total order flow, we expect that the dealer price would increase by the amount of that increase times $\lambda = 0.5 \times (30/5000)^{0.5} = 0.03873$. For example, how would informed order flow and the market maker price level change if uninformed order level were actually 40 shares rather than zero? First, since the informed trader does not know about uninformed orders, his net demand (supply) is unaffected. However, the market maker sets her price based on the sum of total order flow as follows:

$$p = 50 + \frac{1}{2}\sqrt{30/5000}(-64.5497 + 40) = 49.049 = 47.50 + 0.03873 \times 40 = 49.049$$

Since the market maker was unable to distinguish between the informed and uninformed orders, the impact of the informed sell orders was muted somewhat by the noise trader buy orders and the market maker set a market price that reduced slippage from its earlier level above. This camouflage enabled the informed trader to obtain a better price for the sale of his shares.

Kyle: Empirical Evidence

Significant evidence from stock markets, recent and not so recent, supports many predictions of the Kyle model of strategic informed trading. For example, Koudijs (2015), in a simple setting involving trading of English securities in 18th century Amsterdam and London markets, was able to identify the impact of private information. These primitive markets are ideal for examining the applicability of Kyle's model because of the much-delayed arrival of inside information from London to Amsterdam by boats delivering mail packets containing private correspondence and newsletters with public information. This combination of slow information transmission and the absence of insider trading regulation enabled insiders to strategically exploit their superior information, and enabled Koudijs to examine how prices incorporated this private information through the trading process. In particular, insiders would strategically time their more aggressive trading based on the anticipated arrival of the next vessel from London in a manner consistent with Kyle's predictions.

In a similar vein, though based on much later trading scenarios, Bolandnazar, Jackson, Jiang, and Mitts (2020) conducted a "natural experiment" that examined how accidental early releases of security disclosures by the SEC that informed certain traders ultimately incorporated new information into security prices, producing an empirical work that complimented the theoretical work of Kyle (1985). The dates, times, and content of these accidental private releases as well as their later public releases were all known to the researchers. Noise trading and informational advantage impacted aggressive insider trading much as Kyle's model predicted, with less insider trading intensity in the presence of less noise trading and increasing in the insider's relative information advantage. Among the important findings of both Koudijs and Bolandnazar et al. is that insiders trade strategically—for example, when anticipating longer delays to public revelation of private information, they trade more patiently, and otherwise, much as the Kyle model would predict.

6.5 ADVERSE SELECTION, DEALER INVENTORIES, AND THE SPREAD

Walrasian markets assume perfect and frictionless competition where supply of and demand for securities are simultaneously apparent. Market microstructure is, in large part, concerned with how imperfect competition, frictions, and the nonsynchronous arrival of supply and demand impact markets. Walrasian markets are perfectly efficient; market microstructure is concerned with inefficiencies and their impact on the market. In security markets, imperfect competition, bid–offer imbalances, and frictions often reveal themselves in bid–offer spreads. Here, we are concerned with the determinants of the bid–offer spread. First, the evolution of prices through time should provide insight as to what affects the spread. Consider, for example, frictions such as transaction costs and taxes. If these were the only factors affecting the spread, we should expect that in the absence of new information, execution prices would tend to bounce between bid and ask prices. For example, after a trade at the bid price, the next trade would be either at the same price or at the offer. After a trade at the offer, the next trade would be either at the

same price or the bid. That is, price changes would tend to be either zero or at the opposite side of the spread. Thus, execution prices will tend to be either unchanged or to change in the opposite direction from the most recent transaction. Thus, transaction costs tend to induce negative serial correlation in asset prices, particularly in the absence of new information.

On the other hand, asymmetric information tends to produce positive serial correlation in asset prices. Now, suppose that asymmetric information is the only source of the spread, such that transaction prices reflect information communicated by transactions. Transactions executed at bid prices would cause permanent drops in prices to reflect negative information and transactions executed at offer prices would cause permanent increases in prices to reflect positive information. However, the extent to which information arrives randomly will induce a random element in price changes. Thus, if price changes are solely a function of random news arrival, price changes will be random. If the distribution of information is asymmetric, prices will exhibit positive serial correlation as informed traders communicate their information through their trading activity (recall that this is what informed traders try to avoid in the Kyle model). Thus, the extent to which information distribution is asymmetric will affect the serial correlation of asset prices. Increased informational asymmetries imply positively correlated price changes from one transaction to the next.

Inventory costs (such as unsystematic risk caused by the dealer's inability to diversify) will tend to cause negative serial correlations in price quotes. Transactions at the bid will tend to cause risk averse dealers to reduce their bid quotes as they become more reluctant not to overdiversify their inventories. Similarly, transactions at the ask will tend to cause dealers to raise their quotes as they become more reluctant not to underdiversify their inventories. Thus, inventory costs and transaction costs will tend to lead towards negative serial correlation in security prices. As we noted above, asymmetric information availability will lead towards positive serial correlation in security prices as transactions communicate new information.

The Demsetz Immediacy Argument

As we discussed earlier, trader bids and offers do not arrive to markets simultaneously. Order imbalances impose waits on impatient traders requiring immediacy. The costs of providing immediacy to liquidity traders include order processing costs (transaction costs), information and adverse selection costs, inventory holding costs, costs of absorbing inventory risks, and costs of providing trading options. The bid–offer spread provides the dealer compensation for assuming these costs on behalf of the market. Each of these costs is examined in the following sections. But first, in the Demsetz (1968) analysis, buyers and sellers of a security are each of two types, one of which who wants an immediate transaction and a second who wants a transaction, but can wait. Buy and sell orders arrive to the market in a nonsynchronous fashion, causing order imbalances. An imbalance of traders demanding an immediate trade causes price slippage and the less patient trader to pay for immediacy. The greater the costs of trading, and the greater will be the desire for immediacy, the greater will be the market spread.

Glosten and Milgrom Information Asymmetry Model

The Glosten and Milgrom (1985) adverse selection model assumes that dealer spreads are based on the likelihood π (π is a probability, not a profit level in this scenario) that an informed trader (one who knows the value of the asset) will trade. Trades arrive to the market maker, each with some random chance of originating from either an informed or uninformed trader. Suppose that the asset can take on one of two prices, a high price P_H and a low price P_L, each with probability 1/2. The informed trader originates the trade with probability π and knows which price the asset will take on. Liquidity or uninformed traders will not pay more than P_H for the asset and they will not sell for less than P_L. Thus, the informed trader will sell only when the asset is worth P_L and will buy only when the asset is worth P_H. Risk-neutral liquidity traders value the asset at $(P_H + P_L)/2$. When setting his quotes, the dealer needs to account for the probability that an informed trader will transact at his quote, and will set his bid price P_b and ask price P_a as follows:

$$P_b = \pi P_L + (1 - \pi)(P_H + P_L)/2$$
$$P_a = \pi P_H + (1 - \pi)(P_H + P_L)/2$$

The spread is simply the difference between the ask and bid prices:

$$P_a - P_b = \pi(P_H - P_L)$$

Thus, in the single-period Glosten and Milgrom model, the spread is a function of the likelihood that there exists an informed trader in the market and uncertainty in the value of the traded asset. The greater the uncertainty in the value of the traded asset as reflected by $(P_H - P_L)$, and the greater the probability π that a trade has originated with an informed trader, the greater will be the spread. Increased uncertainty and increased information asymmetries lead to increased spreads.

Clearly, traders, including market makers (dealers), will seek to avoid trading with an informed trader and have an interest in knowing the probability that their counterparties in trade will be informed. How might a trader estimate the probability that his counterparty in trade is informed? Suppose that a trader can estimate the risk of a security in a dealer market, and based on his risk estimate, the trader believes that the price of a stock will fall between 20 and 24, with potential prices being uniformly distributed in this range. The trader believes that informed traders will know with certainty the value of the stock. Further suppose that the dealer's bid-offer quotations are 22 and 23, respectively. The trader relies on his risk estimates as well as the dealer's quotes (spread) to determine the probability that an informed trader is active in this market. Based on these quotes provided by the dealer, what is the probability that a given trade will have originated with an informed trader? Since $P_a - P_b = \pi(P_H - P_L)$, $\pi = (P_a - P_b)/(P_H - P_L) = (23\text{-}22)/(24\text{-}20) = .25$. Thus, the trader estimates a 25% probability that his trading counterparty is informed. Essentially, the trader relies on the dealer's knowledge of market participants as reflected through the dealer spread. As the dealer spread widens, the trader associates a higher probability that he would trade with an informed trader. The trader can use this probability to determine whether he should trade at this time. Glosten and Milgrom (1985) extend their modeling over series of sequential trades and more sophisticated scenarios, as do Easley and O'Hara (1987 and 1992).

6.6 RISK, RISK AVERSION, AND DEALER SPREADS

The Stoll Inventory Model

A risk averse dealer needs to maintain inventories in assets in which he makes a market to sell to investors as well as cash to purchase assets from investors. First, we will discuss determinants of dealer behavior in the absence of dealer inventory. Suppose that a dealer currently without inventory in a particular asset P trades so as to maximize the expected utility level that he associates with his uncertain level of wealth W.[2] We assume here that the dealer's wealth level is subject to some normally distributed security return represented by r whose expected value, for sake of simplicity, is assumed to be zero. This security produces an uncertain profit equal to rP and a variance of profits equal to σ^2.[3] The dealer is willing to quote a bid price P_b to purchase this security whose consensus value or consensus price is P. This consensus price can be considered to be the true value estimate for the security. Our problem here is to determine the maximum price that a risk averse dealer would be willing to bid; in the absence of inventory, we show that the level of this bid is a function of the dealer's level of risk aversion. First, we note that the maximum bid that the dealer is willing to quote would be that which equates the expected utility associated with his current uncertain level of wealth after purchasing the risky security with the level of utility he would realize if he opted not to purchase the risky security and its associated risk:

$$E[U(W - P_b + (1 + r)P)] = U(W)$$

Rewriting to adjust expected utility for the expected trading profit against the bid and return over time on the asset consensus value:

$$E(U(W + (P - P_b) + rP)] = U(W)$$

Thus, the dealer's expected utility after the security purchase is a function of his current level of wealth W, his bid price P_b, and his uncertain return. If the security is not purchased, the dealer's utility will be a function of his current wealth level. Our problem is to solve this equality for P_b. We start by performing a Taylor series expansion:

$$E[U(W) + (P - P_b + rP)U'(W)) + 1/2(P - P_b + rP)^2(U''(W)) + \ldots] = U(W)$$

Since $E[r] = 0$, $E[rP]U'(W)$ can be dropped from the equality. In addition, we note that $\sigma^2 = E[(P - P_b + rP)^2]$:

$$E[U(W) + (P - P_b)U'(W)) + 1/2(\sigma)^2(U''(W)) + \ldots] = U(W)$$

We will approximate by dropping all of the left-hand-side higher-order terms not explicitly stated in the above equality, which is quite reasonable if we are willing to assume that

[2]See Stoll (1978). In this discussion, we will assume that the cumulative consensus value of shares traded is P.

[3]Assume that P represents the total investment into asset P. We will not concern ourselves with individual share prices and numbers of shares; we will use P to reflect investment into the tradable asset.

the risk is normally distributed, meaning, for example, that $E[r^3]$ will equal 0. Our pricing equation simplifies as follows:

$$(P - P_b)(U'(W)) \approx -1/2(\sigma)^2 U''(W)$$
$$(P - P_b) \approx -1/2(\sigma)^2 U''(W)/U'(W)$$
$$(P - P_b) \approx 1/2(\sigma)^2 ARA$$

where $-U''(W)/U'(W)$ is the dealer's *Arrow-Pratt absolute risk aversion coefficient* (ARA).[4] The ARA indicates the dealer's aversion ($U''(W) < 0$) to a given risk, based on his utility of wealth function $U(W)$ and his current level of wealth. As the dealer's level of risk aversion increases, the absolute value of $U''(W)$ increases relative to $U'(W)$ since risk aversion is presumed to result from diminishing marginal utility to additional units of wealth. Thus, the greater the dealer's risk aversion, or the greater the uncertainty associated with the asset, the greater will be the discount associated with his bid.

Now, suppose that the dealer maintains an initial inventory W_p of securities constituting his initial wealth level W_p. The dealer is free to borrow and lend at the riskless rate r_f, for sake of simplicity, assumed to be zero. This initial inventory of securities W_p is an optimal portfolio that maximizes the dealer's utility associated with his uncertain level of wealth W subject to his initial wealth constraint:

$$E[U(W_p(1 + r_p) + (P - P_b) + rP)] = U(W_p)$$

where r is the uncertain return on the tradable asset i and r_p is the uncertain return on the optimal portfolio W_p. As we did above, we will expand the left side of this equation around $E[U(W_p)]$, and assuming that $E[r] = E[r_p] = 0$, and after dropping higher-order terms, we have:

$$E[U(W_p) + (P - P_b)(U'(W_p)) + 1/2(\sigma^2)U''(W_p) + 2 \times 1/2 \times E[rP \times r_p W_p] \times U''(W_p)] \approx U(W_p)$$

The covariance of cash flows between profits on the optimal portfolio and the tradable asset i is $2 \times \frac{1}{2} \times E[r_P \times r_p W_p] = \sigma_{i,P}$. A larger dealer inventory of security i in the initially optimal portfolio W_p implies a larger covariance $\sigma_{i,P}$ between i and W_p. Simplifying and solving for $(P - P_b)$, we find that the dealer's bidding discount from the consensus price is:

$$(P - P_b) = \sigma_{i,p} ARA + 1/2(\sigma)^2 ARA$$

Again, the discount $(P - P_b)$ implied by the dealer's bid increases as the dealer's level of risk aversion increases and as the asset risk increases. But, with non-zero inventory levels, the dealer's situation changes, and the bid discount is affected by the covariance of his inventory level and the performance of security i. Notice that a higher level of $\sigma_{i,P}$ implies a higher level of covariance between security i and the dealer's optimal portfolio W_p that

[4]ARA measures how marginal utility changes with respect to a change in wealth relative to how utility changes with respect to a change in wealth. This measure, which is negative for risk averse investors, indicates how much an investor dislikes risk.

includes his inventory of security i. The greater the dealer's inventory level, the greater will be his risk as implied by $\sigma_{i,P}$ along with its associated loss of diversification, and the less he will be willing to bid for an additional unit of security i.

In addition to increasing his bid discount when his level of risk aversion increases and/or when the risk of security i increases, the dealer will increase his bid discount as his inventory of securities increases and as the covariance between his inventory returns and the asset returns increases. Similarly, the premium based the dealer's ask or offer price P_a is obtained as follows:

$$(P_a - P) = \sigma_{i,p} ARA + 1/2(\sigma)^2 ARA$$

Thus, the dealer will increase his offer premium as his inventory of securities decreases and as the covariance between his inventory returns and the asset returns decreases. The bid–offer spread is simply the sum of the bid discount and the offer premium:

$$(P_a - P_b) = (\sigma^2) ARA$$

Thus, the greater the dealer's risk aversion, or the greater the uncertainty associated with the asset, the greater will be the dealer's spread. Interestingly, the dealer's inventory level does not affect the size of the spread, although it will affect each of the two price quotes from which the spread is determined. As the dealer's inventory increases, his reluctance to purchase additional units of the security is offset by his willingness to sell units, simultaneously decreasing both his bid and ask prices by comparable amounts. Consequently, the dealer with non-zero inventory will not place his quotations so that the consensus value of the security is centered within the bid and offer. Instead, the spread will straddle the consensus price, with one quote being closer to the consensus price than the other. Both quotes will decrease as his inventory increases and will increase as his inventory decreases. Other traders can "approximate" the dealer's consensus price (if we were to assume that they not part of the consensus) by knowing the dealer's inventory.

The Copeland and Galai Options Model

Copeland and Galai (1983) reasoned that a bid provides prospective sellers a put on the asset, with the exercise price of the put equal to the bid. For as long as the bid is quoted, other traders (at least the first to exercise) have the right to sell the asset at the quoted price. Similarly, an offer provides a call to other traders. Thus, when the dealer posts both bid and offer quotes, the spread is, in effect, a short *strangle* provided to the market.[5] Increases in underlying asset risk increase both the values of call options and the values of put options. Both legs of this strangle are more valuable when the risk of the underlying security is higher. The value of this short strangle is directly related to the volatility of the asset and inversely related to how quickly it can be revoked or revised.

[5]As we shall discuss in Chapter 7 (see Chapter 1 and the appendix to Chapter 7 for an introduction to options and options pricing), a call is an option to buy the underlying asset and a put is an option to sell the underlying security. A straddle is a combination of put and call options on the same underlying asset; a strangle has different exercise prices.

An options pricing model can be used to value this dealer spread. When the dealer posts quotes, she provides other traders free long and short positions on short-term calls and puts. Thus, dealer posting of quotes is a costly activity. Essentially, a dealer can value each leg of the short strangle as individual options, and add the two to value the short strangle position associated with her quotes. The dealer will then widen the spread to capture this lost value. Ultimately, this "implied premium" associated with this dealer spread is paid by liquidity (uninformed, as in Glosten and Milgrom) traders who trade without information. Informed traders make money at the expense of the dealer, who ultimately earns it back at the expense of uninformed traders. Unlike some of the other models we discussed, the Copeland and Galai model, as presented here, does not have a mechanism to allow trading activity to convey information from informed traders to dealers and uninformed traders. Nonetheless, in the Copeland and Galai option-based model, the spread widens as the uncertainty with respect to the security price increases. Thus, in all three of our spread-based models, Copeland and Galai, Stoll, and Glosten and Milgrom, we see that greater uncertainty increases dealer spreads.

Additional Reading

Chapters 10 to 12 in Harris (2003) offer good introductions to several models presented in this chapter, and O'Hara (1998) offers a somewhat more rigorous introduction. Stoll (2003) and Madhavan (2000) provide excellent reviews of relevant literature as of their publication dates and all four offer broader discussions of market microstructure and related literature. Vayanos and Wang (2013) provide good discussions of both theoretical and empirical issues concerning market liquidity. Amihud et al. (2005) provide a slightly more technical discussion of liquidity and its effects on security prices.

References

Amihud, Y., & Mendelson, H. (1986). Asset pricing and the bid-ask spread. *Journal of Financial Economics*, *17*, 223–249.

Amihud, Y., Mendelson, H., & Pedersen, L. H. (2005). Liquidity and asset prices. *Foundations and Trends in Finance*, *1*(4), 269–364.

Bagehot, W. (1971). The only game in town. *Financial Analysts Journal*, *27*(12–14), 22.

Barber, B., Lee, Y. T., Liu, Y. J., & Odean, T. (2009). Just how much do investors lose from trade? *Review of Financial Studies*, *22*, 151–186.

Barber, B., & Odean, T. (2000). Trading is hazardous to your wealth: The common stock investment performance of individual investors. *Journal of Finance*, *55*, 773–806.

Barber, B., & Odean, T. (2001). Boys will be boys: Gender, overconfidence, and common stock investment. *Quarterly Journal of Economics*, *116*(1), 261–292.

Barber, B., & Odean, T. (2002). On-line investors: Do the slow die first? *Review of Financial Studies*, *15*(2), 455–487.

Barclay, M. J., & Warner, J. B. (1993). Stealth and volatility: Which trades move prices? *Journal of Financial Economics*, *34*, 281–306.

Black, F. (1986). Noise. *Journal of Finance*, *41*(3), 529–543.

Bolandnazar, M., Jackson, R., Jiang, W., & Mitts, J. (2020). Trading against the random expiration of private information: A natural experiment. *Journal of Finance*, *75*(1), 5–44.

Brennan, M. l, Chordia, T., Subrahmanyam, A., & Tong, T. (2012). Sell-order liquidity and the cross-section of expected stock returns. *Journal of Financial Economics*, *105*, 523–541.

Brennan, M., & Subrahmanyam, A. (1996). Market microstructure and asset pricing: On the compensation for illiquidity in stock returns. *Journal of Financial Economics*, *41*, 441–464.

Chan, J., Hong, D., & Subrahmanyam, M. (2008). A tale of two prices: Liquidity and asset prices in multiple markets. *Journal of Banking and Finance*, *32*(6), 947–960.
Chaplinsky, S., & Ramchand, L. (2004). The borrowing costs of international issuers: SEC Rule 144A. *The Journal of Business*, *77*, 1073–1097.
Copeland, T. C., & Galai, D. (1983). Information effects of the bid-ask spread. *Journal of Finance*, *38*, 1457–1469.
Demsetz, H. (1968). The cost of transacting. *Quarterly Journal of Economics*, *82*, 33–53.
Easley, D., & O'Hara, M. (1987). Price, trade size, and information in securities markets. *Journal of Financial Economics*, *19*, 69–90.
Easley, D., & O'Hara, M. (1992). Time and the process of security price adjustment. *Journal of Finance*, *47*, 576–605.
Fama, E. (1965). The behavior of stock market prices. *Journal of Business*, *38*, 34–105.
Figlewski, S. (1979). Market "efficiency" in a market with heterogeneous information. *Journal of Political Economy*, *86*, 581–597.
Friedman, M. (1953). The case for flexible exchange rates. In M. Friedman (Ed.), *Essays in positive economics*. Chicago: University of Chicago Press.
Glosten, L. R., & Milgrom, P. R. (1985). Bid, ask and transactions prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics*, *14*(1), 71–100.
Harris, L. (2003). *Trading and exchanges: Market microstructure for practitioners*. Oxford: Oxford University Press.
Koudijs, P. (2015). Those who know most: Insider trading in 18th century Amsterdam. *The Journal of Political Economy*, *123*, 1356–1409.
Kyle, A. S. (1985). Continuous auctions of insider trading. *Econometrica*, *53*(6), 1315–1336.
Madhavan, A. (2000). Market microstructure: A survey. *Journal of Financial Markets*, *3*, 205–258.
O'Hara, M. (1998). *Market microstructure theory*. New York: Blackwell Publishers.
Rock, K. (1986). Why are new issues underpriced? *Journal of Financial Economics*, *15*, 187–212.
Stoll, H. R. (1978). The supply of dealer services in securities markets. *Journal of Finance*, *33*, 1133–1151.
Stoll, H. R. (2003). Market microstructure. In G. M. Constantinides, M. Harris, & R. Stulz (Eds.), *Handbook of the economics of finance 1A* (pp. 553–604). Amsterdam: Elsevier Science.
Vayanos, D., & Wang J. (2013). Market liquidity: Theory and empirical evidence. In G. M. Constantinides, M. Harris, & R. Stulz (Eds.), *Handbook of the economics of finance 2B* (pp. 1289–1361). Amsterdam: Elsevier Science.
Zarinelli, E., Treccani, M., Farmer, J. D., & Lillo, F. (December 2015). Beyond the square root: evidence for logarithmic dependence of market impact on size and participation rate. *Market Microstructure and Liquidity*, *1*(2), 1137–1156.

6.7 EXERCISES

1. Consider a stock that is traded 4 times per year with an average spread of 1% and paying annual dividends equal to $2, discounted at 3%.
   a. Based on the 1% spread and annual turnover equal to 4, what is the value of the stock?
   b. What would be your answer to part a if the spread were 0.5%?
   c. What would be your answer to part b if the annual turnover were 2?
   d. What is the expected return for the stock in part c?
   e. What are the current bid/ask prices given the scenario in part c? Assume that the stock's value is midway between the bid and ask prices and round to the nearest $0.001.
2. Members of tribal fishing societies often undertake efforts to share information on where the fish are biting. Modern fishermen often purchase information identifying locations of schools of fish obtained from remote sensing satellite data. However, this

information is often kept secret. Why do members of tribal societies share this information and modern fishermen opt not to share?

3. The model of demand in the Kyle (1985) assumes that perfectly informed trader demand x increases linearly in his expected value of the traded stock. This assumption is important to Kyle's results. Demonstrate that perfectly informed trader demand x in the Kyle model increases linearly in the difference between his expected value of the traded stock and the price p set by the dealer. (This difference, $v - p_0$, is the profit per share to the informed trader.)
4. An informed trader has private information that the value of a stock is \$100 per share. Without this information, the variance of payoffs on the stock would be \$60; this is the level of payoff uncertainty faced by uninformed investors. The variance associated with uninformed investor trades is10,000 shares.
   a. If the value of the stock were to be \$100 without the private information, what would be the level of informed trader demand for the stock?
   b. If the value of the stock were to be \$90 without the private information, what would be the level of informed trader demand for the stock?
   c. What will be the informed trader's expected profits from purchasing the number of shares computed in part b?
   d. Suppose that actual uninformed demand for the stock were zero. Based on your computations from part b, at what level would the dealer set the stock price?
5. We saw in Chapter 4 that in a statistical study of stock market data, Barclay and Warner (1993) found that very small trades and the very large trades did not have a significant impact on stock prices. They found that medium-sized trades had the greatest impact on the security process.
   a. Describe how this empirical finding might be consistent with the theoretical results of Kyle (1985).
   b. Describe how this empirical finding might be inconsistent with the theoretical results of Kyle (1985).
   c. Financial blogging has substantially increased the flow of information to investors who would otherwise be uninformed, decreasing investor asymmetries. Would related improvements in the flow of information tend to increase or decrease Kyle's lambda?
6. In the Glosten and Milgrom (1985) model, dealer spreads will tend to be smaller when arrivals of buy and sell orders are balanced. Why do such balanced order arrivals tend to produce smaller dealer spreads?
7. **a.** In the Glosten and Milgrom Information Asymmetry Model, are dealer spreads greater if all traders (excluding the dealer) are fully informed and know what prices will be or if all traders are uninformed and trade on noise? Why?
   b. How is price uncertainty reflected in the Glosten and Milgrom model?
8. Based on his estimates of the risk of a security, a trader believes that the price of a stock traded in a dealer market will fall within a \$3 range, with potential prices being uniformly distributed. The trader knows that there are a substantial number of informed traders trading this stock, and that these informed traders will know with certainty the value of the stock. The dealer's bid−offer quotations are 12.9 and 14.1,

respectively. Based on these quotes provided by the dealer, what is the probability that the trader will estimate that any given trade will have originated with an informed trader?

9. Suppose that a dealer with $20,000 in capital and zero stock inventory has a logarithmic utility of wealth function: $U = \ln(W)$.
   a. Assuming a small risk (the Arrow-Pratt model, as used by Stoll, 1978), what would be his Coefficient of Absolute Risk Aversion (ARA)?
   b. Uncertainty, as measured by the price variance on a given stock is 10,000. The consensus price of this stock is 100. What are the bid price and bid discount for this stock? What are the ask price and ask premium for this stock? What is the spread for this stock?
   c. Now, suppose that the dealer maintains an inventory of shares of this stock so that the covariance between profits on the stock and this optimal initial inventory is $E[r_P \times r_p W_p] = \sigma_{i,P} = 1000$. Calculate the bid discount and price for the stock under this revised circumstance.
   d. What is the dealer spread for part c?
10. Describe how transaction costs will tend to increase inverse correlations between trade-to-trade transaction prices.
11. Suppose that dealer inventory costs were the sole source of the dealer spread. Further suppose that a transaction was executed at the bid. Would subsequent bid and offer quotes tend to be higher or lower? Why? Now, suppose that a transaction was executed at the offer. Would subsequent bid and offer quotes tend to be higher or lower? Why? In either scenario, would transaction prices tend to exhibit positive, negative, or zero autocorrelation? Why?
12. This exercise is related to a trading practice called *quote matching* (see Section 13.4 or Glossary), which occurs when a small trader places an order one uptick (downtick) from that of a large trader so as to profit from the large trader's transaction upward (downward) price pressure, or to use the large trader as a trade counterparty should prices decrease (increase). Suppose that Stock X has just sold for $10.05. An institutional investor places a limit order to purchase 1 million shares at $10.00, which is now the best bid. The best offer is currently $10.10. A trader, who reasons that the stock is equally likely to have an intrinsic value (true worth) of either $10.00 or $10.10 ($E[V] = \$10.05$), places a limit order to buy 200,000 shares of the stock for $10.01.
   a. Describe how the trader is attempting to exploit the options provided by the spread.
   b. How might the institutional investor be losing wealth as a result of his limit order?
13. Based on models described in this chapter, why does the dealer spread widen as the uncertainty associated with the relevant security increases?
14. a. Suppose that stock price changes in a given market from transaction to transaction are clearly and statistically inversely correlated, bouncing between prices in a definable range. Is it more likely that transaction prices are influenced by asymmetric information availability, random arrival of information, or frictions such as transactions costs?
   b. Suppose that stock price changes in a given market from transaction to transaction are clearly uncorrelated, randomly fluctuating over time. Is it more likely that

transaction prices are governed by asymmetric information availability, random arrival of information (that is, perfectly or nearly perfectly efficient markets), or frictions such as transactions costs?

c. Suppose that stock price changes in a given market from transaction to transaction are clearly and statistically positively correlated and are more likely to move in a given direction displaying momentum. Is it more likely that transaction prices are governed by asymmetric information availability, random arrival of information, or frictions such as transactions costs?

APPENDIX

6.A

6.A.1 THE KYLE ADVERSE SELECTION MODEL

This appendix provides for a somewhat more technical introduction to the Kyle (1985) model than is offered in Section 6.4.

Consider a one-time-period single auction model involving an asset that will pay in one time period $v \sim N(p_0; \Sigma_0)$ that is, the asset's future liquidation payoff v is normally distributed with mean value p_0 and variance Σ_0. Thus, p_0 is the unconditional expected value of the asset. Variance, Σ_0, can be interpreted to be the amount of uncertainty that the informed trader's perfect information resolves, so that Σ_0 reflects the informed trader's value advantage. There are three trader types—a single informed trader with perfect information, on the asset payoff many uninformed noise traders, and a single dealer or market maker who sets the market price of the asset and acts as an intermediary in all trades. All are risk neutral, there is no spread, and money has no time value. Market makers and noise traders seek to learn from the actions of the informed trader, who seeks to disguise himself and his special information in a batch market (markets accumulate orders before the dealer clears them). Thus, the informed trader seeks to determine x, an appropriate share purchase (sale) volume that maximizes his trading profits based on his superior information: $\pi = E[(v—p)x|v]$ where p is the market price of the asset. That is, the informed trader seeks to maximize his expected trading profits given his perfect knowledge of v and the price p that he pays (or receives) for x shares of the stock. The market maker will observe total share purchases $X = x+u$, where u reflects noise trader transactions, increasing the price of shares p as X increases. The market maker cannot distinguish between informed trader demand x and noise trader demand u, but does correctly observe total demand X. Nonetheless, X will be correlated with x, hence, the informed trader needs to exercise care in deciding on his transactions volume x to protect himself from price slippage (see Chapter 4). Thus, neither the dealer nor the noise traders will know which trades or traders are informed, but they seek to discern informed demand x from the noisy signal X representing total demand.

Noise traders submit market orders for u shares randomly, such that their orders contain no useful information content. In fact, noise trader activity will obscure information provided by the informed trader, which enables informed traders to disguise the information content of their trading from the dealer. Noise traders will demand u shares, where u is distributed normally with mean $E[u] = 0$ (they are as likely to sell as to buy) and variance $\sigma_u^2 : u \sim N(0; \sigma_u^2)$. Thus, σ_u^2 is the variance of uninformed investor demand. Informed traders do not know how many shares uninformed traders will buy or sell, but the informed

trader does know the parameters of the distribution of the demand. Informed and noise traders submit their order quantities x and u to the market maker in a batch market. The market maker observes the net market imbalance (the extent to which X is positive or negative), and sets the price p at which the total order flow $X=x+u$ is executed and clears. Thus, the market maker observes X, and then sets the price as a function of the sum of informed and uninformed investor demand: $p=E[v|x+u]$.

THE INFORMED TRADER'S PROBLEM: PROFIT MAXIMIZATION

Kyle's Bayesian learning model shows that informed investor demand x can be expressed as a simple linear function of v: $x=\alpha+\beta v$, where α and β are simple coefficients whose traits will be discerned and examined shortly. Similarly, the security's price p, set by the market maker or dealer, is assumed to be a simple linear function of demand: $p=\mu+\lambda(x+u)$, where μ and λ are also simple coefficients whose traits will also be examined shortly. Thus, informed investor demand x is a linear function of true security value v and the security price p is a linear function of the sum of informed and uninformed investor demand $X=(x+u)$.

The informed trader's primary problem is to determine the optimal purchase (or sale) quantity x so as to maximize the expected value of his trading profits:

$$max_x E[\pi] = E[(v-p)x|v] = E[(v-\mu-\lambda(x+u))x|v] = E[vx-\mu x-\lambda x^2-\lambda ux)] \tag{6.A.1}$$

To maximize the informed trader's profits, find the derivative of expected trading profits $E[\pi]$ with respect to informed investor demand x and set equal to zero:

$$\frac{\partial E[\pi]}{\partial x} = v-\mu-2\lambda x-\lambda u = 0 \tag{6.A.2}$$

Also recall that that the distribution of u implies that $E[u]=0$ since uninformed investors randomly buy and sell. Next, note that λ, known as Kyle's lambda, must be positive for the second-order condition (the second derivative must be negative ($-2\lambda<0$)) to hold for maximization. We will demonstrate this later. Finally, we rearrange terms to obtain:

$$2\lambda x = v-\mu \tag{6.A.3}$$

$$x = -\frac{\mu}{2\lambda}+\frac{1}{2\lambda}v \tag{6.A.4}$$

which is linear in v as Kyle proposed it would be. Now, we see that our coefficients α and β are simply:

$$\alpha = -\frac{\mu}{2\lambda}, \quad \beta = \frac{1}{2\lambda}$$

Thus, the informed trader linear demand function $x=\alpha+\beta v=-\frac{\mu}{2\lambda}+\frac{v}{2\lambda}$ is set based on the dealer pricing function.

DEALER PRICE SETTING

But, to gain insight into what the informed trader linear demand function implies, we need to better understand our coefficients α and β. To explore this, we will examine the trading of the dealer or market maker who observes total order flow $X=x+u$, and sets a single market clearing price p or $E[v]$ as a function of total demand $p=E[v|x+u]$ where $x=\alpha+\beta v$. Note here that the dealer increases the price of the security as total demand $x+u$ increases. Since v and X are normally distributed, we will apply the projection theorem to p as follows[6]:

$$p = E[v] + \left[\frac{\text{COV}[v, x+u]}{\text{VAR}[x+u]}\right][x+u-E[x+u]] \tag{6.A.5}$$

This dealer pricing function has a straightforward interpretation. First, the sensitivity of the dealer price to total share demand $x+u$ is a function of the covariance between the stock's value and total demand for the shares COV[v, $x+u$]. Thus, if the dealer believes that total demand $x+u$ for the stock increases dramatically with its intrinsic value v (unknown to him), the price that the dealer sets for shares will be very sensitive to total demand. Thus, if the informed trader is known to dominate trading in the marketplace, the dealer will set the price of the security mostly or entirely as a function of total demand for the security. This sensitivity to total demand will diminish as uninformed demand volatility increases. As total demand deviates more from expected demand, the price of shares will increase.

INFORMED TRADER DEMAND AND DEALER PRICE ADJUSTMENT

Since $x=\alpha+\beta v$, and Σ_0 is the variance of or uncertainty associated with asset payoffs v, the variance of informed trader demand VAR[x] will equal $\beta^2\sum_0$. Since noise trader demand is uncorrelated to informed trader demand, this means that $\text{VAR}[x+u] = \beta^2\sum_0 + \sigma_u^2$, which means that we can write the dealer pricing equation as:

$$p = E[v] + \left[\frac{\text{COV}[v, x+u]}{\beta^2\sum_0 + \sigma_u^2}\right][x+u-\alpha-\beta E[v]] \tag{6.A.6}$$

Now, recall that β is the slope of a line plotting a random dependent variable X or ($x+u$) with respect to an independent random variable v. That is, β=COV[v, $x+u$]/VAR[v] =COV[v, $x+u$]/Σ_0, which means that COV[v, $x+u$]=$\beta\Sigma_0$, and we can write this equation as:

$$p = E[v] + \left[\frac{\beta\sum_0}{\beta^2\sum_0 + \sigma_u^2}\right][x+u-\alpha-\beta E[v]] \tag{6.A.7}$$

[6]The projection theorem has many applications in finance, such as to the Capital Asset Pricing Model (CAPM) and Arbitrage Pricing Theory (APT). Many students are first exposed to it in econometrics when seeking an unbiased estimator in an econometrics setting. Generally, the projection theorem states that the relationship between some random dependent variable y and an independent random variable x is

$$y = E[y] + \left[\frac{\text{COV}[y,x]}{\text{VAR}[x]}\right][x - E[x]]$$

Recall that Kyle suggested a linear relationship between the security price and its demand: $p=\mu+\lambda(x+u)$. This implies a slope λ equal to:

$$\lambda = \left[\frac{\beta\sum_0}{\beta^2\sum_0 + \sigma_u^2}\right] \tag{6.A.8}$$

which implies $\mu=p+\lambda(-x-u)$ and

$$\mu = E[v] + \lambda[-\alpha - \beta E[v]] \tag{6.A.9}$$

Next, we will use α and β coefficients from above to demonstrate that $\mu=E[v]$:

$$\mu = E[v] + \lambda\left[\frac{\mu}{2\lambda} - \frac{1}{2\lambda}E[v]\right] = E[v] + \left[\frac{\mu}{2} - \frac{1}{2}E[v]\right] = \frac{1}{2}E[v] + \frac{\mu}{2} = E[v] \tag{6.A.10}$$

Now, we will rewrite λ, substituting in for β:

$$\lambda = \left[\frac{\beta\sum_0}{\beta^2\sum_0 + \sigma_u^2}\right] = \left[\frac{\frac{1}{2\lambda}\sum_0}{\left(\frac{1}{2\lambda}\right)^2\sum_0 + \sigma_u^2}\right] \tag{6.A.11}$$

We will use a bit of algebra to simplify this expression for λ, starting by multiplying λ and the right-hand side of Equation (6.A.11) by the denominator of the right-hand side of the equation:

$$\lambda\left[\left(\frac{1}{2\lambda}\right)^2\sum_0 + \sigma_u^2\right] = \frac{1}{2\lambda}\sum_0 \tag{6.A.12}$$

Next, we will simplify the left-hand side and then multiply both sides by λ and simplify further by subtracting $\frac{1}{4}\sum_0$ from both sides:

$$\left[\left(\frac{1}{4\lambda}\right)\sum_0 + \lambda\sigma_u^2\right] = \frac{1}{2\lambda}\sum_0 \tag{6.A.13}$$

$$\left(\frac{1}{4}\right)\sum_0 + \lambda^2\sigma_u^2 = \frac{1}{2}\sum_0 \tag{6.A.14}$$

$$\lambda^2\sigma_u^2 = \frac{1}{4}\sum_0 \tag{6.A.15}$$

$$\lambda = \sqrt{\frac{1}{4}\frac{\sum_0}{\sigma_u^2}} = \frac{1}{2}\sqrt{\frac{\sum_0}{\sigma_u^2}} \tag{6.A.16}$$

First, we see that it is obvious that λ is positive and that our second-order condition for profit maximization has been fulfilled. More importantly, λ can be taken to be the dealer

price adjustment for total stock demand; that is, λ *can be considered to be the illiquidity adjustment*. The ratio $\sum_0/\sigma_u^2$ is the ratio of informed trader private information resolution to the level of noise trading. Thus, dealer price adjustment is proportional to the square root of this ratio, increasing as private information Σ_0 is increasing and decreasing as noise trading σ_u^2 increases. This means that if the dealer determines that the informed trader resolves a substantial level of risk relative to the amount of noise trading, the level of dealer price adjustment λ will be large. Now, we can write our informed investor demand coefficients α and β as follows:

$$\alpha = -\frac{E[v]}{\sqrt{\frac{\sum_0}{\sigma_u^2}}}, \qquad \beta = \frac{1}{\sqrt{\frac{\sum_0}{\sigma_u^2}}}$$

Now, we can more easily interpret our informed trader demand function:

$$x = \alpha + \beta v = -\frac{E[v]}{\sqrt{\frac{\sum_0}{\sigma_u^2}}} + \frac{v}{\sqrt{\frac{\sum_0}{\sigma_u^2}}} = \frac{(v - E[v])\sigma_u}{\sqrt{\sum_0}} \tag{6.A.17}$$

This result indicates that informed trader transaction sizes will increase as the variance of uninformed noise demand for shares σ_u^2 increases. That is, the informed trader will trade more aggressively as noise trader volume increases. This volatility reflects noise trader demand for (or supply of) shares. Thus, informed traders will want to take larger positions in the stock or sell off more shares as uninformed trader uncertainty increases and their transaction volume increases. This increased noise volume will better enable informed traders to camouflage their information advantage over the dealer. Also note that, even with perfect information and risk neutrality, informed traders will limit their trading activity to camouflage their intentions. In fact, as the informed trader advantage Σ_0 over the market maker increases, informed trader demand for share volume will decline. That is, as the informed trader's information improves relative to the dealer, the informed trader will seek to camouflage his advantage by reducing his trade volume. This is intended to prevent the dealer from observing the increased trade volume or imbalance on buy (sell) transactions and raising the share price to balance the orders. Here, the informed trader will earn his profits by maintaining more profit on a per share basis rather than on transaction volume.

Recall from Equation (6.A.7) that the market maker sets the price at p, which we will rewrite using the result of Equation (6.A.16):

$$p = E[v] + \left[\frac{\beta\sum_0}{\beta^2\sum_0 + \sigma_u^2}\right][x + u - \alpha - \beta E[v]] = E\left[v\middle|x + u\right]. \tag{6.A.18}$$

$$p = \mu + \lambda(x + u) = E[v] + \frac{1}{2}\sqrt{\frac{\sum_0}{\sigma_u^2}}(x + u) \tag{6.A.19}$$

First, note that the dealer price p is the security's expected value $E[v]$ conditioned on total demand $[x+u]$ for the security. We see here that the dealer price is linear in total demand for the security and linear in the ratio of informed trader value uncertainty to uninformed or noise trader uncertainty. Higher noise or uninformed trader uncertainty reduces the security price unless total demand $(x+u)$ is negative. The opposite is true for value or cash flow uncertainty, which increases the informed trader's informational advantage. The market maker sets the price at p, such that the informed trader buys (sells) whenever $v > E[v]$ $(v < E[v])$, and buys (sells) more aggressively as this difference increases. Note that some of these implications might be clarified with the following informed trader profit function[7]:

$$E[\pi] = E[[v - E[v]]x] = E\left[[v - E[v]]\left[\frac{v}{\sqrt{\frac{\sum_0}{\sigma_u^2}}} - \frac{E[v]}{\sqrt{\frac{\sum_0}{\sigma_u^2}}}\right]\right] = \frac{\sum_0}{\sqrt{\frac{\sum_0}{\sigma_u^2}}} = \sqrt{\sum_0 \sigma_u^2} = \sqrt{\sum_0}\sigma_u \tag{6.A.20}$$

We see here that expected informed trader profits are linear and increasing in the quality of their information advantage $\sqrt{\sum_0}$ and the demand uncertainty σ_u of noise traders. A larger value Σ_0 implies a greater deviation in the security value v (known by the informed trader) from its expected value $E[v]$ (estimated by the dealer). A larger value for Σ_0 implies a larger information advantage to the informed trader. Greater dealer price uncertainty (which arises from the combination of uninformed demand variability and value uncertainty) increases informed trader trading profits. Increased uninformed trader uncertainty and its associated increase in transactions σ_u mean that the informed trader is better able to disguise from the market maker his information advantage Σ_0 and trading activity through the increased noise trader volume σ_u. This ability to camouflage means that the informed trader can trade more aggressively, taking larger positions (x or $-x$) and profits in the stock without accurately revealing his transactions to the market maker. While the market maker does observe order imbalances X, he cannot distinguish between informed trader demand x and noise trader demand u. The market maker responds by setting a competitive price p such that informed traders earn their profits indirectly from noise traders. The market maker loses on trades with informed traders and earns profits on trades with noise traders. The market maker merely earns a competitive profit as long as informed traders successfully camouflage their intentions at the ultimate expense of noise traders. Kyle (1985) continues his model in a multiperiod framework to demonstrate that analogous implications arise in sequential and continuous auctions.

[7]Note that by definition of variance Σ_0, $E[[v - E[v]][v - E[v]]]^2 = \Sigma_0$.

CHAPTER

7

Random Walks, Risk, and Arbitrage

7.1 MARKET EFFICIENCY AND RANDOM WALKS

Market efficiency occurs when prices reflect all available information. Price changes in an efficient market occur when new information becomes available. Since information dissemination (news) arrives randomly, security price changes might be expected to occur randomly. Thus, an efficient market leads to random security price changes. We will discuss market efficiency extensively in Chapter 12; we will introduce sequences of random variables and random walks here.

Random Walks and Martingales

Why should a trader know about random walks and martingales? First, consider the well-known investment company advisory: "Past performance is not indicative of future results."[1] Past performance can be evaluated based on the current price relative to one or more historical prices. If the advisory is true, then future performance is based on a future price and the current price; past prices are irrelevant. In other words, prices follow a *Markov process*, also known as a *random walk*. A random walk is a stochastic process (random sequence) in which a state probability at time t depends only on its immediately prior state and not on the remainder of its history.[2] A stock's price history has no effect on its future if it follows a random walk. If stock prices do not follow random walks, stock markets cannot be efficient.

[1]There are numerous variations of this advisory.

[2]Pearson (1905) described the optimal search process for finding a drunk left in the middle of a field. Left to stagger one step at a time in an entirely unpredictable fashion, she is more likely to be found where she was left than in any other position on the field.

Financial Trading and Investing
DOI: https://doi.org/10.1016/B978-0-323-90955-6.00007-0

© 2023 Elsevier Inc. All rights reserved.

More generally, a *stochastic process* is a sequence of random variables X_t defined on a common probability space $(\Omega,\mathcal{F},\mathbb{P})$ and indexed by time t.[3] In other words, a stochastic process is a random series of values X_t sequenced over time. The values of X_t as t varies define one particular sample path of the process associated with a particular state or outcome from Ω. A *discrete time process* is defined for a countable set of time periods. A *continuous time process* that is defined over an interval of the real number line consisting of an infinite number of time periods. The *state space* X is the set of all possible values of the stochastic process $\{X_t\}$. The state space can be discrete (countable) or continuous. For example, if a stock price changes in increments of \$0.01, the state space is said to be discrete. The state space for prices is continuous if prices can assume any (positive) real value.

As might be inferred from above, a *discrete Markov process* or a *discrete random walk* is a noncontinuous stochastic process in which a state probability at time t depends only on its immediately prior state at time $t-1$:

$$P(X_t|X_{t-1}, X_{t-2}, \ldots, X_0) = P(X_t|X_{t-1})$$

This means that the probability of a given value for X at time t given the historical value of X one period earlier, two periods earlier, and so on all the back to the start of the process is equal to the probability of that given value for X at time t given the value of X one period earlier. Thus, given the process history X_0, X_1, …, X_{t-1} from its start at time 0 to time $t-1$, the probability that that the process will be in state X_t at time t depends only on its present state X_{t-1}. This means that a Markov Process has no memory. Only the current state is relevant to predicting its future.

A *discrete martingale process* is a stochastic process X_t in which $\mathrm{E}[X_t|X_0, X_1, \ldots, X_{t-1}] = X_{t-1}$ for all $t = 1, 2, \ldots$. Thus, a martingale is a process whose future variations have no specific direction that can be discerned from the process history $(X_0, X_1, \ldots, X_{t-1})$. A martingale is said to be a "fair game" because it will not exhibit consistent trends either up or down. A *submartingale* is a stochastic process X_t in which $\mathrm{E}[X_t|X_0, X_1, X_2, \ldots, X_{t-1}] \geq X_{t-1}$. A submartingale will tend either to trend upward over time or is a martingale. A supermartingale will tend to trend downward over time or is a martingale. Stock prices are often modeled as submartingales because they tend to trend upward due to time value of money and investor risk aversion.

Brownian Motion Processes

One particular version of a continuous time-space random walk is a standard *Brownian motion process* Z_t, also known as a Wiener process. A process Z_t is a standard Brownian motion process if:

1. Changes in Z_t over time are independent over disjoint (non-overlapping) intervals of time; that is, $\mathrm{COV}(Z_s - Z_\tau, Z_u - Z_v,) = 0$ when $s > \tau > u > v$.

[3]Mathematicians characterize a probability space as consisting of three types of elements: a sample space Ω of all potential outcomes (e.g., a particular future state of the world), a filtration $\mathcal{F}$, roughly characterized as the information set or history required for valuation (e.g., the security's price history), and an associated probability measure $\mathbb{P}$ that maps subsets from Ω to [0,1]. Readers are likely to encounter this sort of somewhat formal notation in some finance books (See Knopf and Teall, 2015 for more complete explanations), but understanding these more technical mathematical definitions is not essential for the remainder of this text.

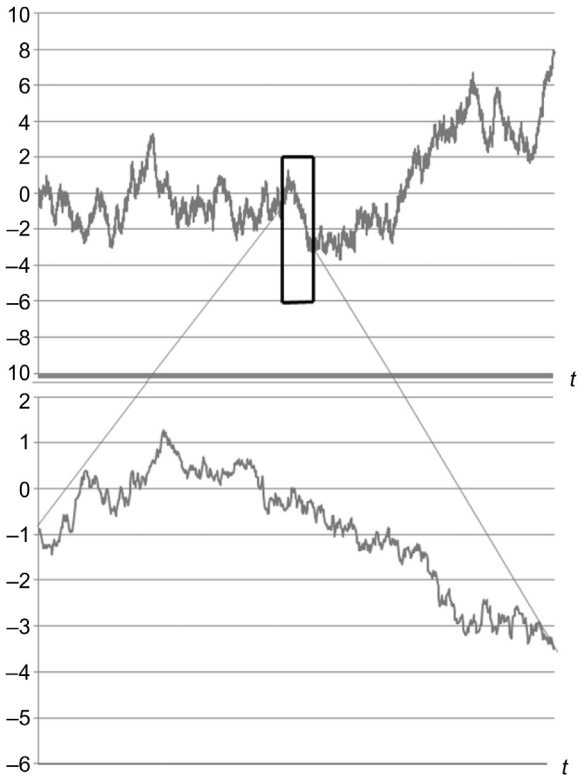


FIGURE 7.1 Brownian motion: a fractal.

2. Changes in Z_t are normally distributed with $E[Z_t - Z_\tau] = 0$ (standard Brownian motion is also a martingale) and $E[(Z_t - Z_\tau)^2] = t - \tau$ for $t > \tau$. Thus, $(Z_t - Z_\tau) \sim N(0, t-\tau)$.
3. Z_t is a continuous function of t.
4. The process begins at zero, $Z_0 = 0$.

Brownian motion has a number of interesting traits. First, it is continuous everywhere and differentiable nowhere (it does not smooth as time intervals decrease) under Newtonian calculus. We see in Figure 7.1 that Brownian motion is a *fractal*, meaning that regardless of the length of the observation time period, the process remains a Brownian motion, and can be defined equivalently by a simple change in the time scale. Consider the Brownian motion process represented by the top graph in Figure 7.1. If a short segment is cut out and magnified as in the bottom graph in Figure 7.1, the graphs look very similar to one another in that they don't smooth and the variance is proportional to time. Further magnifications of cutouts would continue to result in the same phenomenon. In addition, once a Brownian motion hits a given value, it will return to that value infinitely often. More generally, we will scale the variance by a multiple σ^2 so that the change in the process X_t from time t to time $t + \Delta t$ takes the form:

$$\Delta X_t = \sigma \Delta Z_t = \sigma(Z_{t+\Delta t} - Z_t) \sim N(0, \sigma^2 \Delta t)$$

Consider a more general Brownian motion process $S_t = S_0 + \sigma Z_t$. The graph of a particular possible path of S_t over time t is similar to the graph of standard Brownian motion. It is continuous and nondifferentiable (angular movements). The difference is that it can start (time 0) at $S_0 \neq 0$ rather than at the origin 0 and its variance can differ from 1. S_t has a normal distribution with mean S_0 and variance $\sigma^2 t$ ($S_t \sim N(S_0,\sigma^2 t)$) since:

$$\frac{S_t - S_0}{\sigma\sqrt{t}} = \frac{Z_t}{\sqrt{t}} = Z$$

where $Z \sim$ N(0,1). Note that $\Delta S_t = S_{t+\Delta t} - S_t \sim$ N(0,$\sigma^2 \Delta t$).

We can further generalize this Brownian motion process, to make it more appropriate for modeling stock prices S_t. we will generalize further on the Brownian motion process to allow for a trend or drift a in the process:

$$dS_t = adt + \sigma dZ_t$$

where dS_t is the infinitesimal change in the stock price at time t, a represents the drift tendency in the value of S_t, and dZ_t is the infinitesimal change in the standard Brownian motion process. Because prices of many securities such as stocks tend to have a predictable drift component in addition to randomness, this generalized Brownian motion processes might be more practical for modeling purposes than standard Brownian motion, which only includes a random element. We can write a similar process to model stock returns as a Brownian motion with drift:

$$dS_t/S_t = \mu dt + \sigma dZ_t$$

The instantaneous change in the stock price is dS_t, the instantaneous stock return dS_t/S_t trends as per the drift term, μ, which represents the instantaneous expected return for the stock, and σ is the instantaneous stock return standard deviation. The corresponding process for the stock price S_t itself is called geometric Brownian motion (or a geometric Wiener) process. The process allows for compounding of returns as stock prices grow or change. It is the most common process used to model stock prices S_t. This geometric Brownian motion process is particularly useful for modeling the relationships between stocks and options and other derivatives.

7.2 RISK

Definitions of risk vary based on context. Measuring risk can be rather difficult when it is difficult to define risk or to capture all of its characteristics. One perspective on the risk of an investment is that it is simply the uncertainty associated with investment returns or cash flows. However, uncertainty can be a complex quality. Analysts often attempt to quantify risk with absolute measures such as variance or relative risk measures such as beta. Although return variance is a quite simple mathematical concept

with many desirable characteristics, its estimation is hampered by the lack of ideal data. Suppose, for example, that we wish to estimate the risk or variance associated with a stock's returns over the next year. Consider the following discrete expression for ex-ante or forecast variance σ_F^2 that considers all potential return outcomes R_i and associated probabilities P_i:

$$\sigma_F^2 = \sum_{i=1}^{n} (R_i - E[R_i])^2 \, P_i \tag{7.1}$$

While this expression for ex-ante variance is by definition correct, its computation requires that we identify all potential returns for the security (which might range from minus infinity to positive infinity) along with their associated probabilities. This might be practical when the list of potential returns is small or when we can ascertain a specific return generating process (such as Brownian motion). However, associating probabilities with these returns can be a particularly difficult problem. For example, what is the probability that the return for a given stock will range between 5% and 6%? In many instances, we will be forced to either make probability assignments of a somewhat subjective nature or define a joint return and probability generating process for the security. Availability of historical price data can make risk estimation based on historical variances more practical.[4]

Historical Volatility Indicators

Because it is frequently difficult to estimate the inputs necessary to estimate security ex-ante variance, analysts often use the volatility of ex-post or historical returns as a surrogate for ex-ante risk:

$$\sigma_H^2 = \sum_{t=1}^{n} \frac{(R_t - \overline{R}_t)^2}{n-1} \tag{7.2}$$

where R_t represents the return realized during historical period t: ($R_t = [(P_t - P_{t-1})/P_{t-1}]$) in this n time period framework. Table 7.1 presents sample monthly historical price data for GM stock along with monthly returns computed from these prices. We see below that the traditional sample monthly variance estimator for this stock based on these 23 returns equals 0.012: $\sigma_H^2 = \sum_{t=1}^{n} (R_t - \overline{R}_t)^2/(n-1) = 0.012$. If we were to assume that returns follow a Brownian motion process (or at least assume that stock returns are uncorrelated over time), the annualized variance would be $0.144 = 0.012 \times 12$. Use of the traditional sample estimator to forecast variance requires the assumption that stock return variances are constant over time, or more specifically, that historical return variance is an appropriate indicator of future return variance. While this can often be a reasonable assumption, firm risk conditions can change and it is well documented that price volatility does fluctuate over time (see, for example, Officer, 1971). In addition, note that the sample variance estimator rather than the

[4]See Section A.1 in the appendix at the end of the text for a review of elementary statistics and Appendix 7.A.1 at the end of this chapter for a discussion of estimating risk on a spreadsheet.

TABLE 7.1 Traditional Sample Estimators for GM Stock, October 2000 to September 2002

Date	Close	R_t	Date	Close	R_t
3-Sep-02	38.9	−0.187212704	4-Sep-01	41.66	−0.216475456
1-Aug-02	47.86	0.028141783	1-Aug-01	53.17	−0.132343342
1-Jul-02	46.55	−0.12909261	2-Jul-01	61.28	−0.011612903
3-Jun-02	53.45	−0.13998391	1-Jun-01	62	0.130767828
1-May-02	62.15	−0.024026382	1-May-01	54.83	0.047573558
1-Apr-02	63.68	0.061156474	2-Apr-01	52.34	0.05716017
1-Mar-02	60.01	0.141091462	1-Mar-01	49.51	−0.027690495
1-Feb-02	52.59	0.046567164	1-Feb-01	50.92	0.001967729
2-Jan-02	50.25	0.052135678	2-Jan-01	50.82	0.054138146
3-Dec-01	47.76	−0.022113022	1-Dec-00	48.21	0.029028815
1-Nov-01	48.84	0.217044605	1-Nov-00	46.85	−0.195984211
1-Oct-01	40.13	−0.036725876	2-Oct-00	58.27	N/A

population estimator is proposed in Equation (7.2). This difference becomes more significant with smaller samples.

Using Equation (7.2) to estimate security variance requires that the analyst choose a sample series of prices (and dividends, if relevant) at n regular intervals from which to compute returns. Two problems arise in this process:

1. Which prices should be selected and at what intervals?
2. How many prices should be selected?

First, since each of the major stock markets tend to close at regular times on a daily basis, closing prices will usually reflect reasonably comparable intervals, whether selected on a daily, weekly, monthly, annual, or other basis. Those prices closer to the date of computation will probably better reflect security risk (e.g., price volatility of a security 30 years ago hardly seems relevant today). On the other hand, longer-term returns, such as those computed on a monthly or annual basis, might more closely follow a normal distribution than returns computed on a daily or shorter-term basis. This is a highly desirable quality, since many of the statistical estimation procedures used by analysts assume normal (or lognormal) distribution of inputs; the characteristics of most non-normal distributions are not well known.

Generally speaking, more prices or data points used in the computation process will increase the statistical significance of variance estimates. However, this leads to a dilemma: more data points, particularly pertaining to longer-term returns, will require prices from the more distant, but less relevant past. On the other hand, shorter estimation intervals may result in non-normal return distributions as well as autocorrelation issues.

Hence, the analyst must balance the needs for a large sample to ensure statistical significance, recent data for relevance, and longer-term data for independence and normality of distribution. These conflicting needs call for compromise. One convention that has developed over the years both in academia and in the industry is based on computations of five years of monthly returns.

Nonetheless, numerous difficulties still remain with this estimation procedure. For example, as we discussed above, variances are not necessarily stable over time. In our numerical illustration, higher-volatility periods are clustered together as are lower volatility periods, in a manner similar to actual return variances. Second, returns themselves might not be independently distributed. In our numerical illustration, returns are somewhat inversely correlated, potentially leading to significant differences in our variance estimates. Thus, both problems arise in our numerical illustration data in Table 7.1. In addition, nontrading may omit returns data for computations, creating more problems with historical variance estimations. Finally, the traditional historical volatility estimator can involve collecting large amounts of data and tedious calculations.

Extreme Value Estimators

Two difficulties associated with the traditional sample estimator procedure, time required for computation and arbitrary selection of returns from which to compute volatilities might be dealt with by using extreme value estimators. Extreme value estimators are based on high and low values (and sometimes other parameters) realized by the security's price over a given time period.

For example, consider the Parkinson extreme value estimator (Parkinson, 1980). This estimating procedure is based on the assumption that underlying stock returns are lognormally distributed without drift. Given this distribution assumption, the underlying stock's realized high and low prices over a given period provide information regarding the stock's variance. Thus, if we are willing to assume that the return distribution is to be the same during the future period, Parkinson's estimate for the underlying stock return variance is determined as follows:

$$\sigma_p^2 = 0.361 \cdot \left[\ln\left(\frac{HI}{LO}\right)\right]^2 \tag{7.3}$$

where *HI* designates the stock's realized high price for the given period and *LO* designates the low price over the same period. The high and low prices for GM in October 2002 were 47.51 and 38.11 for October 2002, the most recent month listed in Table 7.1. The extreme value estimator for this month would be calculated as follows:

$$\sigma_p^2 = 0.361 \cdot \left[\ln\left(\frac{HI}{LO}\right)\right]^2 = 0.361 \cdot \left[\ln\left(\frac{47.51}{38.11}\right)\right]^2 = 0.017546$$

Alternatively, we could obtain an average monthly variance for the 24-month period covered in the table (ignoring the possibility of intraday highs and lows that differ from

closing prices).[5] Since our price data cover 24 months, we will divide this figure by 24 to obtain our monthly variance. The Parkinson measure results in a monthly variance estimate equal to 0.007414 for GM over the period October 2000 to September 2002 in Table 7.1:

$$\sigma_p^2 = 0.361 \cdot \left[\ln\left(\frac{HI}{LO}\right)\right]^2 / m = 0.361 \cdot \left[\ln\left(\frac{63.68}{38.9}\right)\right]^2 / 24 = 0.007414$$

Clearly, the Parkinson variance estimate will be easy to obtain when periodic high and low prices for a stock are regularly published as they are for the NYSE and many other stocks. Furthermore, the econometric efficiency of the Parkinson procedure is several times higher than the traditional sample estimation procedure. Using the extreme value estimator might be as simple as inserting into Equation (7.3) the 52-week high and low prices. Note that the Parkinson estimate is significantly smaller than the monthly historical estimate. This difference draws largely from the negative autocorrelation (stock price changes are inversely correlated from one time period to the next) in the returns series.

Implied Volatilities[6]

A problem shared by both the traditional sample estimating procedures and the extreme value estimators is that they require the assumption of stable variance estimates over time; more specifically, that future variances equal or can be estimated from historical variances. A third procedure first suggested by Latane and Rendleman (1976) is based on market prices of options that might be used to imply variance estimates. For example, the Black-Scholes option pricing model and its extensions provide an excellent means to estimate underlying stock variances if call prices are known. Essentially, this procedure determines market estimates for underlying stock variance based on known market prices for options on the underlying securities. Consider our stock example from Table 7.1 on September 3, 2002 when GM stock closed at \$38.9. Suppose that a one-month ($T = 1$) call on GM stock with a striking price equal to \$40 was at the same time trading for $c_0 = \$1.16$:

$$T = 1 \qquad r_f = 0.005 \qquad c_0 = \$1.16 \qquad X = \$40 \qquad S_0 = \$38.9$$

[5]Accuracy of the Parkinson measure can be improved if the sample period can be subdivided into n equal sub-periods such that variance is estimated as follows*:

$$\sigma_p^2 = \frac{0.361}{n} \cdot \left[\sum_{t=1}^{n} \ln\left(\frac{HI_t}{LO_t}\right)\right]^2$$

This is the more general form of the Parkinson estimator. We should obtain an intraday highs and lows for each of the 24 months, and then square the sum of their logs. The constant, 0.361, is the normal density function constant, $1/\sqrt{2\pi}$.

[6]This section assumes a background in or a basic understanding of the Black-Scholes options pricing model. If necessary, see Section 7.A.2 in this chapter's appendix for a review of options and the Black-Scholes option pricing model and Section 7.A.3 for additional discussion and examples of estimating implied volatilities.

where r_f equals the monthly riskless return rate and X is the option striking price. If investors use the Black-Scholes option pricing model to value calls, the following must hold:

$$\begin{aligned} 1.16 &= 38.9 \cdot N(d_1) - 40e^{-r_f T} \cdot N(d_2) \\ d_1 &= \frac{\ln\left(\frac{38.9}{40}\right) + (0.005 + 0.5\sigma^2) \cdot 1}{\sigma\sqrt{1}} \\ d_2 &= d_1 - \sigma\sqrt{1} \end{aligned} \tag{7.4}$$

We find that this system of equations holds when $\sigma^2 = 0.01$. Thus, the market prices this call as though it expects that the variance of anticipated monthly returns for the underlying stock is 0.01. The annualized variance would be 0.12 (12 times the monthly variance).

Unfortunately, the system of equations required to obtain an implied variance has no closed form solution. That is, we will be unable to solve Equation (7.4) explicitly for variance; we must search, iterate, and substitute for a solution. You can substitute trial values for σ^2 until you find one that solves the system. A significant amount of time can be saved by using one of several well-known numerical search procedures such as the method of bisection or the Newton-Raphson method.[7]

Table 7.2 presents, describes, and compares basic risk measures discussed in this chapter. There is no consensus as to which measure is clearly best, although some measures

TABLE 7.2 Basic Risk Measures

Measure	Best Used When
Ex-ante measure based on probabilities	1. Ex-ante or future-oriented measure is needed such as when: a. The asset's historical volatility does not properly indicate its future risk. b. The asset's risk characteristics have recently changed. c. The asset has no price or returns history. 2. All potential future return or cash flow outcomes can be specified. 3. Probabilities can be associated with each potential return or cash flow outcome. 4. Instead of (2) & (3), there is a specific return generating process with known parameters.
Traditional sample estimator	1. Variances are expected to be constant between historical and future time periods. 2. There are an appropriate number of sampling intervals where: a. More periods increase statistical significance. b. More periods increase reliance on older, less relevant historical data. 3. Appropriate interval lengths can be determined; longer periods approach normality.
Parkinson extreme value estimator	1. The computationally simplest measure based on a minimum of data is desired. 2. Asset returns are lognormally distributed without drift. 3. Historical volatility is a good indicator of future risk.
Implied volatility: analytical procedures	1. Option prices on relevant underlying asset are readily available. 2. Option pricing model assumptions hold in the relevant market. 3. Can be used when historical volatility does not indicate future risk. 4. User is able to use the appropriate analytical procedures. 5. The market can be assumed capable of assessing risk.

[7]These and other numerical procedures for estimating implied volatilities are discussed in Section 7.A.3 in the Appendix to this chapter. More discussion of the Black-Scholes and other option pricing models is available in Chapter 9.

perform better under some circumstances. Risk measures need to be evaluated on the basis of the availability of appropriate data and appropriateness of underlying assumptions. In addition, all of the measures that we have discussed in this chapter are incomplete in that they do not account for factors such as skewness (unequal concentrations of potential outcomes unequally distributed below and above the mean) and fat tails (kurtosis or high probabilities of extreme events). No single risk measure can serve adequately to capture all sources of security risk. Hence, the trader or securities analyst needs to use judgment in the application and interpretation of risk measures.

7.3 ARBITRAGE

Arbitrage, perhaps the single most important pricing tool in modern finance, is defined as the simultaneous purchase and sale of assets or portfolios yielding identical cash flows. Assets generating identical cash flows should be worth the same amount. This is known as the *Law of One Price*. If assets generating identical cash flows sell at different prices, opportunities exist to create a profit by buying the cheaper asset and selling the more expensive asset. The ability to realize a profit from this type of transaction is known as an *arbitrage opportunity*. Rational investors in such a scenario will seek to purchase the underpriced asset, financing its purchase by simultaneously selling the overpriced asset. The *arbitrageur* will execute such arbitrage transactions, continuing to earn arbitrage profits in increasing quantities until the arbitrage opportunity is eliminated. If markets are competitive, the arbitrageur's purchases of the underpriced asset will bid its price up while the arbitrageur's selling transactions will force down the price of the overpriced asset. These arbitrage transactions should continue until no assets are over- or underpriced. Hence, arbitrageurs should force assets that produce identical cash flow structures to have identical prices.

Classic arbitrage is the simultaneous purchase and sale of the same asset at a profit. For example, if gold is selling in London markets for $1200 per ounce and in New York markets for $1220 per ounce, a classic arbitrage opportunity exists. An investor could purchase gold in London for $1200 per ounce and simultaneously sell it in New York for $1220. This results in a $20 profit per "round-trip" transaction. The transactions involve no risk since both the selling and purchase prices are known and are executed simultaneously. Furthermore, no initial net investment is required because the transactions offset each other; the proceeds of the sale are used to finance the purchase. Thus, if a classic arbitrage opportunity exists, an investor will have the opportunity to make a riskless profit without investing any of his own money. If the laws of supply and demand are not impeded by market inefficiencies, investors will flock to exploit this opportunity. Their buying pressure in London markets will force the London price to rise; their selling pressure in New York markets will force the New York price down. Buying and selling pressure will persist until the prices in the two markets are equal. Thus, significant classic arbitrage opportunities are not likely to persist long in unimpeded free markets. Perhaps an even simpler form of arbitrage opportunity exists in a *crossed market*, where a bid exceeds an offer price. This

can occur when a slow trader's quote is not withdrawn quickly enough, enabling a prospective buyer to purchase a security at the lower offer price and simultaneously sell it at the higher bid price.

More generally, arbitrage refers to the near simultaneous purchase and sale of portfolios generating similar cash flow structures. For example, the cash flow structure of a long position in a forward contract can be replicated by a portfolio consisting of a long position in a call and short position in a put. Although the contracts in the options portfolio are different from the forward contract, the anticipated cash flows are identical. Thus, if the portfolio produces a lower initial cash flow, it can be sold (shorted) while a long position is taken in the forward contract. These positions in sum comprise an arbitrage portfolio. Future cash flows are offset because they are identical, and net sales proceeds from the options portfolio will exceed those of the futures contract. Similarly, arbitrage can refer to offsetting trades of securities whose payoff structures are strongly correlated. For example, the cash flow structure of an automobile industry stock can be nearly replicated with the cash flow structure of a portfolio of other securities designed to be strongly correlated with the stock. When prices of the portfolio and the stock diverge, arbitrage opportunities might arise. Sometimes, these opportunities are referred to as *quasi-arbitrage* in that they are not necessarily completely riskless or that their transactions are not necessarily executed simultaneously.

The principle of arbitrage is the foundation underlying relative securities valuation. That is, we are able to price securities relative to one another or relative to replicating portfolios when arbitrageurs are able to exploit violations of the *law of one price*. The no-arbitrage price of a security is the value of the portfolio constructed to replicate the security. When the law of one price does not hold, one (or both) of the following will hold:

1. There exist opportunities to secure riskless arbitrage opportunities by buying underpriced assets while selling overpriced assets.
2. There exists some sort of market imperfection such as high transactions preventing arbitrageurs from exploiting arbitrage opportunities.

Pairs Trading and Stat-Arb

As with other arbitrage strategies, *pairs trading* involves the simultaneous purchase and sale of similar securities. Pairs trading involves taking offsetting positions in two different stocks (perhaps options or index contracts) with strong returns correlations, one long and one short, such that gains in one position are expected to more than offset losses in the other position. One simple strategy illustration might involve taking a long position in GM stock based on recent price decreases along with a short position in Ford based on recent price increases. If the investor generally expects that automobile stock prices ought to be strongly correlated, and should have been performing similarly, she might expect that future short term performance of the underperforming stock will exceed the performance of the stock with the stronger current price level. Pairs trading is essentially an arbitrage strategy anticipating that the deviation of a recent pricing relation between two securities

is only temporary. Holding periods for most pairs trading strategies tend to be quite short, ranging from seconds to days. Pairs traders typically focus either on the ratio between prices of two securities or the difference between their prices. One cannot expect to consistently earn high returns based on such strategies, so that many pairs traders execute many such trades over many different pairs of stocks with the expectation that their strategies will be correct more often than they are wrong. Furthermore, many pairs traders will seek to hold "portfolios of pairs" whose expected returns will tend to be uncorrelated with those of the market as a whole. Such "beta-neutral" portfolios might be expected to be insulated from large losses related to stock market swings.

In a manner similar to pairs trading, *stat-arb* (statistical arbitrage) strategies seek to exploit mispricing opportunities while minimizing risk. Stat-arb strategies focus on statistical or historical relationships among securities and seek to exploit price divergences as portfolios of securities diverge from "normal" pricing relationships. Khandani and Lo (2007) refer to stat-arb as a set of "highly technical short-term mean-reversion strategies involving large numbers of securities (hundreds to thousands, depending on the amount of risk capital), very short holding periods (measured in seconds to days), and substantial computational, trading, and IT infrastructure." There is no single stat-arb strategy; there are many types, all of which involve large sets of securities whose statistical price relationships relative to each other diverge from what the investor expects.

7.4 LIMITS TO ARBITRAGE

Arbitrage opportunities are the traders' "pot of gold at the end of the rainbow." Once found, they are free sources of money. In theory, arbitrage is riskless because asset cash flows are known to be offsetting in the future and because initial transactions are executed simultaneously at known prices. Unfortunately, the reality is often quite different. For example, *implementation risk* arises because transactions might not be executed or be executed at prices that differ from what the arbitrageur anticipated. Perhaps even worse, one leg of the arbitrage transaction might fail to execute, leaving the arbitrageur unhedged in an undesirable portfolio. Arbitrageur profit opportunities should cause markets to respond more efficiently to new information and mispricing of securities. However, sometimes arbitrageurs simply don't fully understand or appreciate the models that they trade on. This misunderstanding is called *model risk*. In many instances, model risk arises from a simple failure to appreciate the differences between model assumptions and reality. Despite potential profit opportunities, some markets are slow to react to arbitrage opportunities and some do not react at all. In fact, some markets may move opposite to the arbitrageur's expectations, at least in the short run. This might render traders unable to survive in the long run. This risk is called *basis risk*. If trading is difficult or expensive, perhaps due to high transaction costs, price adjustments to arbitrage may be delayed, prevented, or again, move further in the unanticipated direction. Furthermore, the arbitrageur is subject to margin calls, forced sell-outs and forced buy-ins, potentially shortening her time horizon to recover from the unanticipated price change. Basis risk is mitigated if the arbitrageur is able to count on a known terminal liquidation position (such as with short-term options, futures, and Treasury bills) and maintain her portfolio hedge until that date.

Thus, short-lived securities such as options tend to have smaller basis risk than long-lived securities such as common stock. An additional source of arbitrage risk, particularly when instruments are not exchange-traded, is *counterparty risk*, which is the potential that a trade counterparty fails to fulfill his side of a transaction.

De Long et al. (1990) argue that the short investment horizons of arbitrageurs can prevent them from taking advantage of errors or noise in market prices. If it appears that price deviations from fundamental values might persist, arbitrageurs might avoid potentially profitable opportunities, causing even greater risk and deviations from fundamental values. This might enable noise traders to earn even higher returns from their pricing errors, causing pricing bubbles to form. For example, consider the case of Long Term Capital Management (LTCM), which was founded in the early 1990s by John Meriwether, formerly head of the fixed income unit of Salomon Brothers, Inc. He brought in famed academics Myron Scholes and Robert Merton, developers of some of the most important derivatives and arbitrage strategies of the time, and the fund focused much of its trading activity on derivatives and fixed income. LTCM realized enormous returns and growth until 1998, when it lost over $4 billion in one spectacularly disastrous quarter, forcing the fund into liquidation and threatening contagion throughout financial markets and potential market meltdown. In many instances as we will discuss shortly, LTCM took appropriate positions in arbitrage portfolios, but the fund was so highly leveraged that it could not withstand short-term market moves against their positions.

Among the most important strategies to LTCM was to exploit *convergence trades*, which often involved taking long positions in cheaper foreign debt and short positions in more expensive U.S. Treasury securities. The foreign debt instruments were generally cheaper because they were not as liquid; thus they had higher yields. However, holding positions over the long term meant that the cheaper foreign debt would eventually rise in prices relative to the more expensive U.S. Treasury securities. LTCM, with its less than $5 billion in equity capital, controlled over $100 billion in assets and approximately $1.25 trillion notional in derivative contracts. Unfortunately for LTCM, the Southeast Asian financial crisis in 1997−1998 followed by the Russian financial crisis of 1998 caused a "flight to liquidity," meaning in this case that investors sold non-U.S. debt, forcing down non-U.S. debt prices, and purchased U.S. Treasury securities, forcing U.S. debt prices up. LTCM's equity capital was wiped out within days, and the fund was bailed out by a consortium of financial institutions under the supervision of the U.S. Federal Reserve System. While LTCM's strategy may well have been correct had it been able to survive for an extended period of time, the fund was too highly leveraged to survive extreme short-term volatility.

One of LTCM's equity market arbitrage ventures involved the Royal Dutch/Shell Group, a dual-listed company independently incorporated in the Netherlands and the UK. This arrangement originated from a 1907 alliance agreement between Royal Dutch and Shell Transport in which the two companies agreed to merge their interests on a 60/40 basis. Royal Dutch trades primarily in the United States and the Netherlands while Shell trades primarily in London. After adjusting for foreign exchange rates, shares of the two firms' stock should trade at a 1.5:1 ratio. However, this anticipated trading ratio has deviated by more than 35%, well beyond levels of tax differentials and transaction costs. In the summer of 1997, Royal Dutch traded at an 8% to 10% premium over its 1.5 expected level relative to Shell. To exploit this differential, LTCM had taken significant arbitrage positions on the two

stocks. As the differential widened to 20% in 1998, LTCM increased its positions until financial distress caused by trading activities elsewhere in the fund (related to the economic crisis in Russia) forced the firm to liquidate. The positions taken in Royal Dutch/Shell by LTCM were ultimately proven correct, but not until 2001 when the fund was no longer in existence. Rosenthal and Young (1990) argue that significant mispricing in dual-listed companies has prevailed over long periods of time without satisfactory explanations. Nevertheless, arbitrage can sometimes take significant amounts of time to equalize prices, causing arbitrageurs to maintain positions longer than they are able.

In another well-publicized failure of arbitrage, dating from a March 2000 equity carve-out, 3Com spun off its Palm division, a maker of handheld computers. 3Com retained 95% of the shares of Palm and announced that each 3Com shareholder would ultimately receive 1.5 shares of Palm for each share of 3Com. The remaining 5% of Palm shares were issued at $38 per share, increasing to $165 by its first day of trading before closing at $95.06 (see Lamont, 2002). Remember that ownership of one 3Com share implied ownership of 1.5 shares of Palm stock. The stocks of the two companies should have moved in tandem, but on the date of the IPO, 3Com actually decreased by 21% to $81.81 as Palm increased. This $81.81 is substantially less than the $142.59 price implied by the 1.5 shares of Palm stock due to each 3Com shareholder ($1.5 \times \$95.06 = \$142.59$), implying that the remainder of 3Com, on a per share basis, was worth negative $60.78. This negative stub value (the whole is worth less than the sum of the parts; in particular, the parent and the subsidiary are worth less than the subsidiary alone) seems particularly unlikely, since 3Com had about $10 per share in cash and marketable securities alone. In other words, what happened wasn't rational, given the numbers, or just could not have been sustainable had investors been able to arbitrage. On the other hand, prospective arbitrageurs found themselves unable to short sell shares because the two stocks were under different national regulatory authorities and the outcome of the Palm carve-out was still uncertain. Thus, arbitrage and price correction could not be implemented because the short selling mechanism was not available for the Palm IPO. Therefore, as Cherkes, Jones and Spatt (2013) argue, the apparent mispricing scenario might have been perfectly rational given the prevailing market inefficiencies.

Such negative stub values are not uncommon. For example, in 1923, Benjamin Graham chronicled his purchase of shares of stock in Du Pont, a well-established firm that had negative stub value given its investment in the new company General Motors. Lamont and Thaler (2003) identified five other 1990s technology equity carve-outs with negative stub values: UBID, Retek, PFSWeb, Xpedior, and Stratos Lightwave. Arbitrage in each of these cases was impeded by the inability to short sell. Many research analysts (e.g., Colvino, 2014) argued that between 2014 and 2016, given its investment in Yahoo Japan and Alibaba, the stub value of Yahoo's core business was negative, perhaps as much as negative $40 billion, despite significant short interest in the shares. Mitchell et al. (2002) found 82 similar negative stub value scenarios in U.S. markets between 1985 and 2000. Nevertheless, in most cases, arbitrage was impeded by inability to short sell, high transaction costs, and difficulty in getting reliable price quotes or other information. But, Mitchell et al. found that approximately 30% of negative stub values were never eliminated through arbitrage. Some of the spin-offs failed, and others may have faced this risk. But, even these conditions probably cannot explain particularly large negative stub values.

There are even more obvious limits on many arbitrage opportunities. Restrictions on trading and transaction costs can be major barriers to arbitrage. For example, Han and Wang (2004) found that upper and lower fractional ownership bounds on the holdings of a stock can limit institutional arbitrage and contribute to the momentum of returns for that stock. For example, many countries such as China restrict proportional ownership and even the United States imposes insider trading regulations on major shareholders (e.g., the Williams Amendment). Pontiff (1996) argued that large differences in prices and net asset values in closed end funds result when fund portfolios are more difficult to replicate, when trading costs are high for the stocks in fund portfolios, and when stocks in the funds paid dividends. Thus, difficulties in replicating assets do limit arbitrage opportunities. Similarly, traders know well that Treasury instruments regularly trade at lower yields than nearly identical "off-the-run Treasuries." Violations of the simple put-call parity relation regularly arise when there are restrictions on short-selling underlying securities. Transaction costs often require that the arbitrageur set upper and lower bounds on security purchase and selling prices, essentially determining the interval within which the no-arbitrage security price must fall.

Additional Reading

Pearson (1905) provides a nice, 100-year-old nontechnical discussion of the random walk process. Knopf and Teall (2015) offer a more technical introduction to random walks, martingales, and stochastic processes in finance. Lamont (2002) provides a very readable case study referred to in this chapter concerning the anomalous behavior of the Palm and 3-Com shares following their takeover scenario. The case is important because it depicts the failure of the market to provide rational relative pricing for the two firms. Lamont and Thaler (2003) provide a listing of other firms demonstrating negative stub values. Lowenstein's 2000 book *When Genius Failed: The Rise and Fall of Long-Term Capital Management* is an excellent history of the rise and fall of LTCM.

References

Brenner, M., & Subrahmanyam, M. G. (1988). A simple formula to compute the implied standard deviation. *Financial Analysts Journal*, *5*, 80–83.

Cherkes, M., Jones, C., & Spatt, C. (2013). *A solution to the Palm—3Com spin-off puzzles*. Unpublished working paper. New York: Columbia University.

Colvino, S. (2014). *Yahoo stub value*. Markit Research, September 30, 2014. Available at <https://www.markit.com/Commentary/NewsCommentarieFile?CMSID = aade814d646749c3a7b2701931e606d0> Accessed 27.07.16.

Corrado, C., & Miller, T. W. (1996). A note on a simple, accurate formula to compute implied standard deviations. *Journal of Banking and Finance*, *20*, 595–603.

De Long, J. B., Shleifer, A., Summers, L. H., & Waldman, R. J. (1990). Noise trader risk in financial markets. *Journal of Political Economy*, *98*(4), 703–738.

Han, B., & Wang, Q. (2004). *Institutional investment constraints and stock prices*. Dice Center Working Paper No. 2004-24. Available at: <http://ssrn.com/abstract = 628683> Accessed 12.01.12.

Khandani, A. E., & Lo, A. W. (2007). What happened to the quants in August 2007? *Journal of Investment Management*, *5*(4), 5–54.

Knopf, P., & Teall, J. (2015). *Risk neutral pricing and financial mathematics: A primer*. San Diego: Academic Press.

Lamont, O. (2002). The curious case of Palm and 3 Com. In J. Pickford (Ed.), *Mastering investment* (pp. 261–266). New York: Prentice Hall.

Lamont, O., & Thaler, R. (2003). Can the market add and subtract? Mispricing in tech stock carve-outs. *Journal of Political Economy*, *111*, 227–268.

Latane, H., & Rendleman, R. (1976). Standard deviations of stock price ratios implied by option prices. *Journal of Finance*, *31*, 369–381.
Lowenstein, R. (2000). *When genius failed: The rise and fall of Long-Term Capital Management*. New York: Random House.
Mitchell, M., Pulvino, T., & Stafford, E. (2002). Limited arbitrage in equity markets. *Journal of Finance*, *57*(2), 551–584.
Officer, R.A. (1971). *A time series examination of the market factor of the new york stock exchange* (Ph.D. dissertation). University of Chicago.
Parkinson, M. (1980). The extreme value method for estimating the variance of the rate of return. *Journal of Business*, *57*, 61–65.
Pearson, K. (1905). The problem of the random walk. *Nature*, *72*, 342.
Pontiff, J. (1996). Costly arbitrage: Evidence from closed end funds. *Quarterly Journal of Economics*, *111*, 1135–1152.
Rosenthal, L., & Young, C. (1990). The seemingly anomalous price behavior of Royal Dutch/Shell and Unilever N.V./PLC. *Journal of Financial Economics*, *26*, 123–141.

7.5 EXERCISES

1. Now, let's consider martingale and Markov processes in a standard Blackjack game. Cards are dealt one at a time from a standard 52-card randomly shuffled deck, and points are awarded to the lone recipient based on the number on the card (2 to 10) or 11 if the dealt card is a face card or Ace. Let S_t represent the number of points to be held by the recipient after t cards have been dealt by the dealer. For parts a through c, suppose that the cards have been dealt without replacement. For parts d, e, and f, assume that the cards have been dealt with replacement and that 1 point is awarded if the number on the card is a 2 through a 6, 0 points are awarded if the number on the card is 7, 8, or 9, and -1 point is awarded if the card is a 10, a face card, or an ace. We note that this is the most common point system used by card counters playing blackjack in casinos.
   a. Is this process stochastic?
   b. Is this process Markov? This part of the problem is a little more difficult.
   c. Is this process a submartingale?
   d. Is this process (with replacement) Markov?
   e. Is this process (with replacement) a submartingale?
   f. Is this process (with replacement) a martingale?
2. Let $\{r_t, t \geq 0\}$ (the return on a stock) be an arithmetic Brownian motion.
   a. Suppose that r_t is made up of two components, an instantaneous drift with expected value $\mu = .05$ and a variance $\sigma^2 = .25$. What is the probability that r_5 is between .3 and .5?
   b. Suppose that the price of a stock follows a geometric Brownian motion process. Suppose that the stock's initial value is $S_0 = 1$, and its instantaneous drift r_t has an expected value $\mu = .05$ per year and an annual variance $\sigma^2 = .25$. What is the probability that the stock is worth more than 2 in five years $P[S_5 > 2]$?
   c. Is the return process for this stock a martingale?
3. Suppose a stock price S_t follows a Brownian motion process with an initial price of $S_0 = \$50$ and a variance per time period equal to $\sigma^2 = 4$. What is the probability that the stock price is less than \$56 at time 3?

4. Historical *percentage* stock returns for the Robinson and Boyer Companies are listed in the following chart along with percentage returns on the market portfolio:

Year	Robinson	Boyer	Market
2008	4	19	15
2009	7	4	10
2010	11	−4	3
2011	4	21	12
2012	5	13	9

Calculate the following based on the preceding table:

a. mean historical returns for the two companies and the market portfolio.

b. variances associated with Robinson Company returns and Boyer Company returns as well as returns on the market portfolio.

c. Forecast a variance and a standard deviation of returns for both the Robinson and Boyer Companies based on your calculations in parts a and b.

5. The following table represents outcome numbers, probabilities, and associated returns for Stock A:

Outcome (i)	Return (R_i)	Probability (P_i)
1	05	10
2	15	10
3	05	05
4	15	10
5	15	10
6	10	10
7	15	10
8	05	10
9	15	?
10	10	10

Thus, there are 10 possible return outcomes for Stock A.

a. What is the probability associated with outcome 9?

b. What is the standard deviation of returns associated with Stock A?

6. The Nettles Company management projects an expected return level of 15% for the upcoming year. Assuming that returns are normally distributed with a standard deviation equal to. 1, what is the probability that the actual return level will:

a. fall between 5% and 25%?

b. fall between 15% and 25%?

c. exceed 25%?
d. exceed 30%?

7. What would each of the probabilities in problem 4 be if Nettles Company management were certain enough of its forecast to associate a 5% standard deviation with its returns projection?

8. The following table presents sample daily historical price data for a stock whose returns are given in the third column.

Time	Price_t	Return_t
0	30.000	N.A.
1	30.125	0.00417
2	30.250	0.00415
3	30.125	−0.00413
4	32.000	0.06224
5	34.000	0.06250
6	31.000	−0.08824
7	32.000	0.03226
8	30.500	−0.04688
9	30.750	0.00820
10	30.875	0.00407
11	31.000	0.00405
12	30.875	−0.00403
13	31.000	0.00405
14	31.125	0.00403
15	30.250	−0.02811
16	33.000	0.09091
17	30.000	−0.09091
18	35.125	0.17083
19	33.000	−0.06050
20	32.125	−0.02652
21	32.250	0.00389
22	32.375	0.00388
23	32.125	−0.00772
24	32.250	0.00389
25	34.250	0.06202
26	36.375	0.06204

27	38.500	0.05842
28	34.375	−0.10714
29	33.875	−0.01455
30	33.625	−0.00738

a. Based on a traditional sample estimator, calculate a daily variance estimator for this stock.
b. Assume that returns follow a Brownian motion process (at least that stock returns are uncorrelated over time) and that there are 30 trading days per month. What would be the monthly variance for this stock?
c. What would be the Parkinson extreme value estimated daily returns variance for this stock?

9. The following daily prices were collected for each of three stocks over a 12-day period.

Company X		Company Y		Company Z	
Date	**Price**	**Date**	**Price**	**Date**	**Price**
1/09	50.125	1/09	20.000	1/09	60.375
1/10	50.125	1/10	20.000	1/10	60.500
1/11	50.250	1/11	20.125	1/11	60.250
1/12	50.250	1/12	20.250	1/12	60.125
1/13	50.375	1/13	20.375	1/13	60.000
1/14	50.250	1/14	20.375	1/14	60.125
1/15	52.250	1/15	21.375	1/15	62.625
1/16	52.375	1/16	21.250	1/16	60.750
1/17	52.250	1/17	21.375	1/17	60.750
1/18	52.375	1/18	21.500	1/18	60.875
1/19	52.500	1/19	21.375	1/19	60.875
1/20	52.375	1/20	21.500	1/20	60.875

Based on the data given above, calculate the following:

a. Returns for each day on each of the three stocks. There should be a total of 10 returns for each stock—beginning with the date 1/10.
b. Average daily returns for each of the three stocks.
c. Daily return standard deviations for each of the three stocks.

10. Torre Company stock realized a 52-week high of $50 per share and a 52-week low of $25. What is the Parkinson extreme value estimate for variance for this stock? What would be the corresponding standard deviation estimate?

11. Suppose that there is a six-month call currently trading for $8.20 while its underlying stock is currently trading for $75. Other details for this example are as follows:

$$T = 0.5 \quad r_f = 0.10 \quad c_0 = 8.20 \quad X = 80 \quad S_0 = 75$$

What is the volatility (standard deviation) implied by the market price of this call?

12. Santo Company stock currently trades for $50 per share. The current riskless return rate is 0.06. Under the Black-Scholes framework, what would be the standard deviations implied by six-month (0.5 year) European calls with current market values based on each of the following striking prices?

a. $X = 40$; $c_0 = 11.50$
b. $X = 45$; $c_0 = 8.25$
c. $X = 50$; $c_0 = 4.75$
d. $X = 55$; $c_0 = 2.50$
e. $X = 60$; $c_0 = 1.25$

13. In principle, simultaneous execution of transactions at known prices producing offsetting future cash flows makes arbitrage riskless. However, in practice, even in the simplest scenarios, such arbitrage does involve several risks. What are these risks?

APPENDIX

7.A

7.A.1 RETURN AND RISK SPREADSHEET APPLICATIONS

Table 7.A.1 contains spreadsheet entries for computing stock variances, standard deviations, and covariances. The table lists daily closing prices for Stocks X, Y, and Z from January 9 to January 20 in cells B3:B14, E3:E14, and H3:H14. From these prices, we compute returns in Cells B19:B29, E19:E29, and H19:H29. Variance, standard deviation, and covariance statistics in rows 30 to 38 are computed from formulas displayed in Table 7.A.2.

Formulas for computing returns are given in rows 19 to 29 in Table 7.A.2. Means, variances, standard deviations, covariances, and correlation coefficients are computed in rows 30 to 38. Row 30 computes the arithmetic mean return for each of the three stocks. Table 7.A.2 lists formulas associated with the values in cells A30:H38. The = (Average) function may be typed in directly as listed in Table 7.A.2 row 30 or obtained from the Paste Function button (f_x) menu under the Statistical submenu. Entry instructions are given in the dialogue box obtained when the Average function is selected. The variance formulas in row 31 are based on the sample formula; the variance (P) formulas in row 32 are based on the population formula. Standard deviation sample and population results are given in rows 33 and 34. Covariances and correlation coefficients are given in rows 35 to 38.

7.A.2 A PRIMER ON BLACK-SCHOLES OPTION PRICING

Calls and Puts

First, we will introduce a few option basics. A *stock option* is a legal contract that grants its owner the right (not obligation) to either buy or sell a given stock. There are two types of stock options: puts and calls. A *call* grants its owner to purchase stock (called underlying shares) for a specified exercise price (also known as a striking price or exercise price) on or before the expiration date of the contract. In a sense, a call is similar to a coupon that one might find in a newspaper enabling its owner to, for example, purchase a roll of paper towels for one dollar. If the coupon represents a bargain, it will be exercised and the consumer will purchase the paper towels. If the coupon is not worth exercising, it will simply be allowed to expire. The value of the coupon when exercised would be the amount by which value of the paper towels exceeds one dollar (or zero if the paper towels are worth less than one dollar). Similarly, the value of a call option at exercise equals the difference between the underlying market price of the stock and the exercise price of the call.

TABLE 7.A.1 Stock Prices, Returns, Risk, and Co-Movement

	A	B	C	D	E	F	G	H
1	CORP. X			CORP. Y			CORP. Z	
2	DATE	PRICE		DATE	PRICE		DATE	PRICE
3	9-Jan	50.125		9-Jan	20		9-Jan	60.375
4	10-Jan	50.125		10-Jan	20		10-Jan	60.5
5	11-Jan	50.25		11-Jan	20.125		11-Jan	60.25
6	12-Jan	50.25		12-Jan	20.25		12-Jan	60.125
7	13-Jan	50.375		13-Jan	20.375		13-Jan	60
8	14-Jan	50.25		14-Jan	20.375		14-Jan	60.125
9	15-Jan	52.25		15-Jan	21.375		15-Jan	62.625
10	16-Jan	52.375		16-Jan	21.25		16-Jan	60.75
11	17-Jan	52.25		17-Jan	21.375		17-Jan	60.75
12	18-Jan	52.375		18-Jan	21.5		18-Jan	60.875
13	19-Jan	52.5		19-Jan	21.375		19-Jan	60.875
14	20-Jan	52.375		20-Jan	21.5		20-Jan	60.875
15								

16	CORP. X		-	CORP. Y		-	CORP. Z	
17	DATE	RETURN		DATE	RETURN		DATE	RETURN
18	9-Jan	N/A		9-Jan	N/A		9-Jan	N/A
19	10-Jan	0		10-Jan	0		10-Jan	0.00207
20	11-Jan	0.002494		11-Jan	0.00625		11-Jan	−0.00413
21	12-Jan	0		12-Jan	0.006211		12-Jan	−0.00207
22	13-Jan	0.002488		13-Jan	0.006173		13-Jan	−0.00208
23	14-Jan	−0.00248		14-Jan	0		14-Jan	0.002083
24	15-Jan	0.039801		15-Jan	0.04908		15-Jan	0.04158
25	16-Jan	0.002392		16-Jan	−0.00585		16-Jan	−0.02994
26	17-Jan	−0.00239		17-Jan	0.005882		17-Jan	0
27	18-Jan	0.002392		18-Jan	0.005848		18-Jan	0.002058
28	19-Jan	0.002387		19-Jan	−0.00581		19-Jan	0
29	20-Jan	−0.00238		20-Jan	0.005848		20-Jan	0
30	Mean	0.004064		Mean	0.006694		Mean	0.00087

(*Continued*)

TABLE 7.A.1 (Continued)

16	CORP. X		CORP. Y		CORP. Z	
17	DATE	RETURN	DATE	RETURN	DATE	RETURN
31	Variance	0.000145	Variance	0.00022	Variance	0.000266
32	Variance (P)	0.000132	Variance (P)	0.0002	Variance (P)	0.000241
33	St.D.	0.01204	St.D.	0.014842	St.D.	0.016296
34	St.D. (P)	0.011479	St.D. (P)	0.014151	St.D. (P)	0.015538
35	COV(X,Y) =		0.0001494	COV(Y,Z) =	0.000192	
36	COV(X,Z) =		0.000139			
37	CORR(X,Y) =		0.9196541	CORR(Y,Z) =	0.8733657	
38	CORR(X,Z) =		0.7791748			

Suppose, for example, that there is a call option with an exercise price of $90 on one share of stock. The option expires in one year. This share of stock is expected to be worth either $80 or $120 in one year, but we do not know which at the present time. If the stock were to be worth $80 when the call expires, its owner should decline to exercise the call. It would simply not be practical to use the call to purchase stock for $90 (the exercise price) when it can be purchased in the market for $80. The call would expire worthless in this case. If, instead, the stock were to be worth $120 when the call expires, its owner should exercise the call. Its owner would then be able to pay $90 for a share that has a market value of $120, representing a $30 profit. In this case, the call would be worth $30 when it expires. Let T designate the options term to expiry, S_T the stock value at option expiry, and c_T be the value of the call option at expiry determined as follows:

$$c_T = MAX[0,\ S_T - X] \tag{7.A.2.1}$$

$$\text{When } S_T = 80,\ \ c_T = \text{MAX}[0,\ 80 - 90) = 0$$

$$\text{When } S_T = 120,\ \ c_T = \text{MAX}[0,\ 120 - 90) = 30$$

A put grants its owner the right to sell the underlying stock at a specified exercise price on or before its expiration date. A put contract is similar to an insurance contract. For example, an owner of stock may purchase a put contract ensuring that he can sell his stock for the exercise price given by the put contract. The value of the put when exercised is equal to the amount by which the put exercise price exceeds the underlying stock price (or zero if the put is never exercised).

To continue the above example, suppose that there is a put option with an exercise price of $90 on one share of stock. The put option expires in one year. Again, this share of stock is expected to be worth either $80 or $120 in one year, but we do not know which yet. If the stock were to be worth $80 when the put expires, its owner should exercise the put. In this case, its owner could use the put to sell stock for $90 (the exercise price) when

TABLE 7.A.2 Stock Returns, Risk, and Co-Movement: Formula Entries

	A	B	C	D	E	F	G	H
1		CORP. X			CORP. Y			CORP. Z
2	DATE	RETURN		DATE	RETURN		DATE	RETURN
18	9-Jan	N/A		9-Jan	N/A		9-Jan	N/A
19	10-Jan	= B4/B3-1		10-Jan	= E4/E3-1		10-Jan	= H4/H3-1
20	11-Jan	= B5/B4-1		11-Jan	= E5/E4-1		11-Jan	= H5/H4-1
21	12-Jan	= B6/B5-1		12-Jan	= E6/E5-1		12-Jan	= H6/H5-1
22	13-Jan	= B7/B6-1		13-Jan	= E7/E6-1		13-Jan	= H7/H6-1
23	14-Jan	= B8/B7-1		14-Jan	= E8/E7-1		14-Jan	= H8/H7-1
24	15-Jan	= B9/B8-1		15-Jan	= E9/E8-1		15-Jan	= H9/H8-1
25	16-Jan	= B10/B9-1		16-Jan	= E10/E9-1		16-Jan	= H10/H9-1
26	17-Jan	= B11/B10-1		17-Jan	= E11/E10-1		17-Jan	= H11/H10-1
27	18-Jan	= B12/B11-1		18-Jan	= E12/E11-1		18-Jan	= H12/H11-1
28	19-Jan	= B13/B12-1		19-Jan	= E13/E12-1		19-Jan	= H13/H12-1
29	20-Jan	= B14/B13-1		20-Jan	= E14/E13-1		20-Jan	= H14/H13-1
30	Mean	= AVERAGE (B19:B29)		Mean	= AVERAGE (E19:E29)		Mean	= AVERAGE (H19:H29)
31	Variance	= VAR(B19:B29)		Variance	= VAR (E19:E29)		Variance	= VAR(H19:H29)
32	Variance (P)	= VARP (B19:B29)		Variance (P)	= VARP (E19:E29)		Variance (P)	= VARP(H19:H29)
33	St.D.	= STDEV (B19:B29)		St.D.	= STDEV (E19:E29)		St.D.	= STDEV (H19:H29)
34	St.D. (P)	= STDEVP (B19:B29)		St.D. (P)	= STDEVP (E19:E29)		St.D. (P)	= STDEVP (H19:H29)
35		COV(X,Y) =		= COVAR (B19:B29,E19:E29)	COV(Y,Z) =		= COVAR (E19:E29,H19:H29)	
36		COV(X,Z) =		= COVAR (B19:B29,H19:H29)				
37		CORR(X,Y) =		= CORREL (B19:B29,E19:E29)	CORR(Y,Z) =		= CORREL (E19:E29,H19:H29)	
38		CORR(X,Z) =		= CORREL (B19:B29,H19:H29)				

it can be purchased in the market for $80. The put would be worth $10 in this case. If, instead, the stock were to be worth $120 when the put expires, its owner should not exercise the put. Its owner should not accept $90 for a share that has a market value of $120. In this case, the put would be worth nothing when it expires. Let p_T be the value of the put option at expiry, determined as follows:

$$p_T = \text{MAX}[0,\ X - S_T] \tag{7.A.2.2}$$

$$\text{When } S_T = 80,\ \ p_T = \text{MAX}[0,\ 90 - 80) = 10$$

$$\text{When } S_T = 120,\ \ p_T = \text{MAX}[0,\ 90 - 120) = 0$$

The owner of the option contract may exercise his right to buy or sell; however, he is not obligated to do so. Stock options are simply contracts between two investors issued with the aid of a clearing corporation, exchange, and broker, which ensure that investors honor their obligations to each other. The corporation whose stock options are traded will probably not issue and does not necessarily trade these options.

For each owner of an option contract, there is a seller or "writer" who creates the contract, sells it to a buyer, and must satisfy an obligation to the owner of the option contract. The option writer sells (in the case of a call exercise) or buys (in the case of a put exercise) the stock when the option owner exercises. The owner of a call is likely to profit if the stock underlying the option increases in value sufficiently over the exercise price of the option (he can buy the stock for less than its market value); the owner of a put is likely to profit if the underlying stock declines in value sufficiently below the exercise price (he can sell stock for more than its market value). Since the option owner's right to exercise represents an obligation to the option writer, the option owner's profits are equal to the option writer's losses. Therefore, an option must be purchased from the option writer; the option writer receives a "premium" from the option purchaser for assuming the risk of loss associated with enabling the option owner to exercise.

The Black-Scholes Model

The Black-Scholes option pricing model provides a simple mechanism for valuing calls under certain assumptions (see Chapter 9 for more detail on the Black-Scholes model and its assumptions). If circumstances are appropriate to apply the Black-Scholes model, call options can be valued with the following:

$$c_0 = S_0 N(d_1) - \frac{X}{e^{r_f T}} N(d_2) \tag{7.A.2.3}$$

$$d_1 = \frac{\ln(S_0/X) + (r_f + \sigma^2/2)T}{\sigma\sqrt{T}} \tag{7.A.2.4}$$

$$d_2 = d_1 - \sigma\sqrt{T} \tag{7.A.2.5}$$

where $N(d^*)$ is the cumulative normal distribution function for (d^*). This function might be referred to in a statistics setting as the "z" value for (d^*). From a computational perspective, one would first work through Equation (7.A.2.4) and then Equation (7.A.2.5) before

valuing the call with Equation (7.A.2.3). $N(d_1)$ and $N(d_2)$ are areas under the standard normal distribution curves (z-values). Simply locate the z-value on an appropriate table (see Table 7.A.1 in the text appendix) corresponding to the $N(d_1)$ and $N(d_2)$ values.

Consider the following simple illustration of a Black-Scholes model application: An investor has the opportunity to purchase a six month call option for $7.00 on a stock that is currently selling for $75. The exercise price of the call is $80 and the current riskless rate of return is 10% per annum. The variance of annual returns on the underlying stock is 16%. At its current price of $7.00, does this option represent a good investment? First, we note the model inputs in symbolic form:

$$T = 0.5 \quad r_f = 0.10 \quad \sigma = 0.4 \quad S_0 = 75$$

$$X = 80 \quad \sigma^2 = 0.16 \quad e \approx 2.71828$$

Our first steps are to find d_1 and d_2 from Equations (7.A.2.4) and (7.A.2.5):

$$d_1 = \frac{\ln(75/80) + (0.1 + 0.5 \cdot 0.16) \cdot 0.5}{0.4 \cdot \sqrt{0.5}} = \frac{\ln(0.9375) + 0.09}{0.2828} = 0.09$$

$$d_2 = d_1 - 0.4 \cdot \sqrt{0.5} = 0.09 - 0.2828 = -0.1928$$

Next, by either using a z-table (see Table 7.A.1 in the text Appendix) or by using an appropriate estimation function from a statistics manual, we find normal density functions for d_1 and d_2:

$$N(d_1) = N(0.09) = 0.536 \quad N(d_2) = N(-0.1928) = 0.420$$

Finally, we use $N(d_1)$ and $N(d_1)$ in Equation (7.A.2.3) to value the call:

$$c_0 = 75 \cdot (0.536) - (80 \cdot 0.9512) \cdot (0.420) = 8.23$$

Since the 8.23 estimated value of the call exceeds its 7.00 market price, the call should be purchased.

Appendix Exercises

1. Call and put options with an exercise price of $30 are traded on one share of Company X stock.
   - **a.** What are the value of the call and the put if the stock is worth $33 when the options expire?
   - **b.** What are the value of the call and the put if the stock is worth $22 when the options expire?
   - **c.** What is the value of the call writer's obligation if the stock is worth $33 when the options expire? What is the value of the put writer's obligation if the stock is worth $33 when the options expire?
   - **d.** What is the value of the call writer's obligation if the stock is worth $22 when the options expire? What is the value of the put writer's obligation if the stock is worth $22 when the options expire?

e. Suppose that the purchaser of a call in part a paid \$1.75 for his option. What was his profit on his investment?

f. Suppose that the purchaser of a call in part b paid \$1.75 for his option. What was his profit on his investment?

2. Evaluate calls for each of the following European stock option series:

Option 1	Option 2	Option 3	Option 4
$T = 1$	$T = 1$	$T = 1$	$T = 2$
$S = 30$	$S = 30$	$S = 30$	$S = 30$
$\sigma = 0.3$	$\sigma = 0.3$	$\sigma = 0.5$	$\sigma = 0.3$
r_f=0.06	$r_f = 0.06$	r_f=0.06	$r_f = 0.06$
$X = 25$	$X = 35$	$X = 35$	$X = 35$

Appendix Exercise Solutions

1.

a. $c_T = \$33 - \$30 = \$3$; $p_T = 0$

b. $c_T = 0$; $p_T = \$30 - \$22 = \$8$

c. $c_T = -\$3$; $p_T = 0$

d. $c_T = 0$; $p_T = -\$8$

e. $\$3 - \$1.75 = \$1.25$

f. $\$0 - \$1.75 = -\$1.75$

2. The options are valued with the Black-Scholes model in a step-by-step format in the following table:

	Option 1	Option 2	Option 3	Option 4
$d(1)$	0.957739	−0.163836	0.061699	0.131638
$d(2)$	0.657739	−0.463836	−0.438301	−0.292626
$N[d(1)]$	0.830903	0.434930	0.524599	0.552365
$N[d(2)]$	0.744647	0.321383	0.330584	0.384904
Call	7.395	2.455	4.841	4.623

7.A.3 ESTIMATING IMPLIED BLACK-SCHOLES VARIANCES

Analysts often employ historical return variances to estimate the volatility of securities. However, one cannot always assume that variances will be constant over time or that historical data properly reflects current conditions. An alternative procedure to estimate security variances is based on the assumption that investors price options based on consideration of the underlying stock risk. If the price of the option is taken to be correct, and if the Black-Scholes option pricing model is appropriate for valuing options, then one can infer the underlying stock standard deviation based on the known market price of the

option and the option pricing model. Consider the following example pertaining to a six-month call currently trading for $8.20 and its underlying stock currently trading for $75:

$$T = 0.5 \quad r_f = 0.10 \quad c_0 = 8.20$$
$$X = 80 \quad S_0 = 75$$

If investors have used the Black-Scholes option pricing model to evaluate this call, the following must hold:

$$8.20 = 75 \times N(d_1) - 80 \times e^{-0.01 \times 0.5} \times N(d_2)$$

$$d_1 = \{\ln(75/80) + (0.1 + 0.5\sigma^2) \times 0.5\} \div \sigma\sqrt{0.5}$$

$$d_2 = d_1 - \sigma\sqrt{0.5}$$

Thus, we wish to solve the above system of equations for σ. This is equivalent to solving for the root of:

$$f(\sigma^*) = 0 = 75 \times N(d_1) - 80 \times e^{-0.1 \times 0.5} \times N(d_2) - 820$$

based on equations above for d_1 and d_2. There exists no closed form solution for σ. Thus, we will use the method of bisection to search for a solution. We first arbitrarily select endpoints $b_1 = 0.2$ and $a_1 = 0.5$ such that $f(b_1) < 0$ and $f(a_1) > 0$. Since these endpoints result in $f(\sigma)$ with opposite signs, our first iteration will use $\sigma_1 = 0.5(0.2 + 0.5) = 0.35$. We find that this estimate for sigma results in a value of -1.30009 for $f(\sigma)$. Since this $f(\sigma)$ is negative, we know that σ* is in the segment $b_2 = 0.35$ and $a_2 = 0.5$. We repeat the iteration process, finding after 19 iterations that $\sigma^* = 0.411466$. Table 7.A.3 details the process of iteration.

The Newton-Raphson method can also be used to more efficiently iterate for an implied volatility. Consider the following example where we wish to estimate the volatility implied by a six-month option with an exercise price of $80 currently selling for $8.20. Assume that the underlying stock price is currently $75 and that the riskless return rate is 0.10. We shall solve for the implied standard deviation using the Newton-Raphson method, with an arbitrarily selected initial trial solution of $\sigma_1 = 0.6$. First, we need the derivative of the Black-Scholes model with respect to the underlying stock return standard deviation[8]:

$$\frac{\partial C}{\partial \sigma} = S\sqrt{t}\frac{e^{\frac{-(d_1^2)}{2}}}{\sqrt{2\Pi}} > 0 \quad Vega = \nu \tag{7.A.3.1}$$

We see from Table 7.A.4 that this standard deviation results in a value of $f(\sigma_0) = 3.950117$. Plugging 0.6 into Equation (7.A.3.1) for σ, we find that $f'(\sigma_1) = 20.82508$. Thus, our second trial value for σ is determined by: $\sigma_2 = 0.6 - (3.950117 \div 20.82508) = 0.410319$. This process continues until we converge to a solution of approximately 0.411466. Note that the rate of convergence is much faster by using the Newton-Raphson method than by using the method of bisection.

[8]See Chapter 9 for additional details and discussion concerning vega.

TABLE 7.A.3 Using the Bisection Method to Estimate Implied Volatility

Initial Equation: $SN(d_1) - Xe^{-r_f T}N(d_2)$

$a_1 = 0.5$ $b_1 = 0.2$ $\sigma_1 = 0.35$
$r_f = 0.1$ $S_0 = 75$ $X = 80$
$c_0 = 8.2$ $T = 0.5$

				$d_1(\sigma n)$	$d_2(\sigma n)$	$N(d_1)$	$N(d_2)$	$N(d_1)$	$N(d_2)$	$f(\sigma n)$
$f(a_1) =$		**1.860465**	**0.5**	**0.1356555**	**−0.2178978**	**0.553953**	**0.41375**	**0.553953**	**0.413754**	**1.860465**
$f(b_1) =$		**−4.46788**	**0.2**	**−0.0320922**	**−0.1735135**	**0.487199**	**0.431122**	**0.487199**	**0.431124**	**−4.46788**
n	a_n	b_n	σ_n	$d_1(\sigma_n)$	$d_2(\sigma_n)$	$N(d_1)$	$N(d_2)$	$N(d_1)$	$N(d_2)$	$f(\sigma_n)$
1	0.5	0.2	0.35	0.06499919	−0.1824882	0.525913	0.427597	0.525913	0.4276	−1.29619
2	0.5	0.35	0.425	0.10188237	−0.198638	0.540575	0.42127	0.540575	0.421273	0.284948
3	0.425	0.35	0.3875	0.08394239	−0.1900615	0.533449	0.424628	0.533449	0.424630	−0.50501
4	0.425	0.3875	0.40625	0.09302042	−0.1942417	0.537056	0.42299	0.537056	0.422993	−0.10987
5	0.425	0.40625	0.41562	0.09747658	−0.1964147	0.538826	0.42214	0.538826	0.422143	0.087583
6	0.41562	0.40625	0.41093	0.09525501	−0.1953217	0.537944	0.42256	0.537944	0.422571	−0.01113
7	0.41562	0.41093	0.41328	0.09636739	−0.1958666	0.538386	0.42235	0.538386	0.422357	0.038229
8	0.41328	0.41093	0.41210	0.09581161	−0.1955937	0.538165	0.42246	0.538165	0.422464	0.01355
9	0.41210	0.41093	0.41152	0.09553341	−0.1954576	0.538054	0.42251	0.538054	0.422517	0.00121
10	0.41152	0.41093	0.41123	0.09539424	−0.1953896	0.537999	0.42254	0.537999	0.422544	−0.00496
11	0.41152	0.41123	0.41137	0.09546383	−0.1954236	0.538027	0.42252	0.538027	0.422531	−0.00188
12	0.41152	0.41137	0.41145	0.09549862	−0.1954406	0.538041	0.42252	0.538041	0.422524	−0.00033
13	0.41152	0.41145	0.41148	0.09551602	−0.1954491	0.538048	0.42251	0.538048	0.422521	0.000438
14	0.41148	0.41145	0.41146	0.09550732	−0.1954449	0.538044	0.42251	0.538044	0.422522	0.000053
15	0.41146	0.41145	0.41145	0.09550297	−0.1954427	0.538042	0.42252	0.538042	0.422523	−0.00014
16	0.41146	0.41145	0.41146	0.09550514	−0.1954438	0.538043	0.42252	0.538043	0.422523	−0.00004
17	0.41146	0.41146	0.41146	0.09550623	−0.1954443	0.538044	0.42252	0.538044	0.422523	0.000004
18	0.41146	0.41146	0.41146	0.09550569	−0.1954441	0.538043	0.42252	0.538043	0.422523	−0.00002
19	0.41146	0.41146	0.41146	0.09550596	−0.1954442	0.538044	0.42252	0.538044	0.422523	0.00000
20	0.41146	0.41146	0.41146	0.09550610	−0.1954443	0.538044	0.42252	0.538044	0.422523	0.00000
21	0.41146	0.41146	0.41146	0.09550616	−0.1954443	0.538044	0.42252	0.538044	0.422523	0.00000
22	0.41146	0.41146	0.41146	0.09550613	−0.1954443	0.538044	0.42252	0.538044	0.422523	0.00000
23	0.41146	0.41146	0.41146	0.09550615	−0.1954443	0.538044	0.42252	0.538044	0.422523	0.00000

TABLE 7.A.4 Newton-Raphson Method and Implied Volatilities

n	σ_n	$f'(\sigma_n)$	$d_1(\sigma_n)$	$d_2(\sigma_n)$	$N(d_1)$	$N(d_2)$	$N(d_1)$	$N(d_2)$	$f(\sigma_n)$
1	0.600000	20.82508	0.177864	−0.2463997	0.57058527	0.40268	0.570585	0.402686	3.950117
2	0.410319	21.06193	0.094961	−0.1951785	0.53782718	0.42262	0.537827	0.422627	−0.024150
3	0.411466	21.06084	0.095506	−0.1954443	0.53804365	0.42252	0.538044	0.422523	0.000000
4	0.411466	21.06084	0.095506	−0.1954443	0.53804365	0.42252	0.538044	0.422523	0.000000
5	0.411466	21.06084	0.095506	−0.1954443	0.53804365	0.42252	0.538044	0.422523	0.000000
6	0.411466	21.06084	0.095506	−0.1954443	0.53804365	0.42252	0.538044	0.422523	0.000000

$$\text{Initial Equation: } S_0N(d_1) - Xe^{-r_fT}N(d_2)$$

$$r_f = 0.1 \quad S_0 = 75 \quad X = 80$$

$$c_0 = 8.20 \quad T = 0.5 \quad \sigma_0 = 0.6$$

Simple Closed-Form Procedures

The implied volatilities described above have the desirable characteristic of having an ex-ante orientation. However, their use is somewhat complicated by the frequently time-consuming methodology of iterating for solutions. Spreadsheet file and even computer program solutions packages can sometimes be cumbersome. This section provides two methodologies for obtaining simple closed form solutions for implied underlying asset volatilities.

Brenner and Subrahmanyam (1988) provide a simple formula to estimate an implied standard deviation (or variance) from an option whose striking price equals the current market price of the underlying asset. As the market price differs more from the option striking price, the estimation accuracy of this formula will worsen:

$$\sigma_{BS}^2 = \frac{2\pi c_0^2}{tS_0^2} \tag{7.A.3.2}$$

The accuracy of Equation (7.A.3.2) when the option strike and underlying asset market prices are unequal can be improved with use of a quadratic procedure proposed by Corrado and Miller (1996). Their formula makes use of a second-order approximation of the cumulative normal density function and an approximation for $\ln(S_0/X)$ as $2(S_0 - X)/(S_0 + X)$. Several additional simplifications and approximations lead to the following formula for σ:

$$\sigma_{CM}^2 = \frac{2\pi}{t(S_0+X)^2} \times \left[c_0 - \frac{S_0 - X}{2} + \sqrt{\left(c_0 - \frac{S_0 - X}{2} \right)^2 - \frac{(s_0 - X)^2}{\pi}} \right] \tag{7.A.3.3}$$

While the accuracy of this formula is improved when the market price of the underlying asset is close to the striking price of the option, it is not nearly as sensitive to differences between these prices as is the Brenner-Subrahmanyam formula.

Aggregating Procedures

A difficulty arising with estimating implied variances results from the fact that there will typically be more than one option trading on the same stock. However, what if the short- and long-term uncertainty of a stock differs? Each option's market price will imply its own underlying stock variance, and these variances are likely to differ. Worse, what if the implied volatility of option contracts differs with respect to exercise prices? (This is called a volatility smile.) How might we use this conflicting information to generate the most reliable variance estimate? Each of our implied variance estimates is likely to provide some information, yet has the potential for having measured with error. We can preserve much of the information from each of our estimates and eliminate some of our estimating error if we use for our own implied volatility a value based on an average of all of our estimates. However, because volatility might be expected to vary over time, one should average only those variances implied by options with comparable terms to expiration. The following suggests two methods for averaging implied standard deviation estimates:

1. Simple average: Here, the final standard deviation estimate is simply the mean of the standard deviations implied by the market prices of the calls.
2. Average based on price sensitivities to σ: Calls that are more sensitive to σ as indicated by $\partial c_j/\partial \sigma_j$ are more likely to imply a correct standard deviation estimate. Suppose we have n calls on a stock and each of which implies a stock standard deviation σ_j. Each call price will have a sensitivity to its implied underlying stock standard deviation $\partial c_j/\partial \sigma_j$. The sensitivities can be summed, and a weighted average standard deviation estimate for the underlying stock can be computed from the following weighting scheme:

$$\omega_i = \frac{\dfrac{\partial c_{0,i}}{\partial \sigma_i}}{\displaystyle\sum_{j=1}^{n} \frac{\partial c_{0,j}}{\partial \sigma_j}} \tag{7.A.3.4}$$

where w_i represents the weight for the implied standard deviation estimate for call option i. Thus, the final standard deviation estimate for a given stock k based on all of the implied standard deviations from each of the call prices is:

$$\sigma_k = \sum_{i=1}^{n} \omega_i \sigma_i \tag{7.A.3.5}$$

CHAPTER

8

Arbitrage and Hedging With Fixed Income Instruments and Currencies

8.1 ARBITRAGE WITH RISKLESS BONDS

As we discussed in Chapter 7, traders are very concerned with arbitrage, where riskless arbitrage involves the simultaneous purchase and sale of identical cash flows. The most easily arbitraged financial instruments are often those with guaranteed payments or with payments that are perfectly correlated with other instruments. Consider a set of riskless bonds whose coupon payments and redemption values are guaranteed. Such riskless bonds might be replicated by portfolios of other riskless bonds if their payments are similarly guaranteed and are to be made on the same dates. For example, consider the illustration provided in Table 8.1, which consists of three priced riskless bonds, all making payments on the same dates until they mature. The cash flow structure of any one-, two-, or three-year bond (e.g., Bond D) added to the market can be replicated with some portfolio of bonds A, B, and C, as long as all the cash payments to investors on Bond D are on the same dates as those made by at least one (in this example, two) of the three bonds A, B, and C.[1] For example, assume that there now exists Bond D, a three-year, 20% coupon bond selling in this market for $1360. This bond will make payments of $200 in years 1 and 2 in addition to a $1200 payment in year 3. We will demonstrate that a portfolio of bonds A, B, and C can be assembled to generate the same cash flow series as Bond D.

[1]More precisely, the set of payoff vectors for Bonds A, B, and C must span the 3-dimensional payoff space. More generally, a set of m vectors $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m\}$, where each vector $\mathbf{x}_i$ is an n-dimensional vector in $\mathbb{R}^n$, is said to *span* the n-dimensional vector space $\mathbb{R}^n$ if any vector in $\mathbb{R}^n$ can be expressed as a linear combination of the vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m$. In other words, for every vector $\mathbf{v}$ in $\mathbb{R}^n$, there exist scalars $\alpha_1, \alpha_2, \ldots, \alpha_m$ such that $\mathbf{v} = \alpha_1\mathbf{x}_1 + \alpha_2\mathbf{x}_2 + \cdots + \alpha_m\mathbf{x}_m$. Intuitively for our illustration, this implies that the payoff vector of any Bond D can be replicated by some portfolio (with appropriate weights w_A, w_B, and w_C) of bonds A, B, and C.

Financial Trading and Investing
DOI: https://doi.org/10.1016/B978-0-323-90955-6.00008-2

© 2023 Elsevier Inc. All rights reserved.

TABLE 8.1 Coupon Bonds A, B, and C

Bond	Current Price	Face Value	Coupon Rate	Years to Maturity
A	1000	1000	0.04	2
B	1055.5	1000	0.06	3
C	889	1000	0	3

Thus, as we will demonstrate, Bond D can be replicated by a portfolio of our first three bonds in the following quantities: $b_A = 0$, $b_B = 3\frac{1}{3}$, and $b_C = -2\frac{1}{3}$, which are determined by the following system of equations:

$$\begin{aligned} 200 &= 40b_A + 60b_B + 0b_C \\ 200 &= 1040b_A + 60b_B + 0b_C \\ 1200 &= 1060b_B + 1000b_C \end{aligned}$$

This system sets the cash flows each year from Bond D equal to those of a portfolio comprising specific numbers of Bonds A, B, and C. There are many ways to solve this system for the numbers (b) of Bonds A, B, and C to place in this replicating portfolio. One such way is to start by subtracting the first equation from the second equation, obtaining the following:

$$0 = 1000b_A$$

This is easily solved for $b_A = 0$. Now, substitute $b_A = 0$ into our original second equation:

$$200 = 1040b_A + 60b_B$$

Now, we see that $b_B = 3\frac{1}{3}$. Substitute this result into our original third equation:

$$\begin{aligned} 1200 &= 1060b_B + 1000b_C \\ 1200 &= 3533.33 + 1000b_C \end{aligned}$$

Now, solve this to obtain $b_C = -2\frac{1}{3}$.

Thus, we find from this system that $b_A = 0$, $b_B = 3\frac{1}{3}$, and $b_C = -2\frac{1}{3}$. We determine the value of the portfolio replicating Bond D by weighting their current market prices: $(0 \times \$1000) + (3\frac{1}{3} \times \$1055.5) + (-2\frac{1}{3} \times 889) = \$1444$. Based on the portfolio's price, the value of Bond D is \$1444, although its current market price is \$1360. Thus, a trader can gain an arbitrage profit from the purchase of this bond for \$1360 financed by the sale of the portfolio of Bonds B and C at a price of \$1444. Shorting the portfolio implies selling $3\frac{1}{3}$ Bonds B and purchasing $2\frac{1}{3}$ Bonds C. Our cash flows in years 1, 2, and 3 will be zero, although we receive a positive cash flow now or time zero equal to \$1444 − 1360 = 84. This is a clear arbitrage profit. This arbitrage opportunity will persist until the value of the portfolio equals the value of Bond D.

Using matrix notation from text Appendix A.2, we have the following system to replicate Bond D:

$$\begin{bmatrix} 200 \\ 200 \\ 1200 \end{bmatrix} = \begin{bmatrix} 40 & 60 & 0 \\ 1040 & 60 & 0 \\ 0 & 1060 & 1000 \end{bmatrix} \begin{bmatrix} b_A \\ b_B \\ b_C \end{bmatrix}$$
$$\mathbf{cf_d} = \mathbf{CF} \quad \mathbf{b}$$

To solve this system we first invert matrix **CF**, then use it to premultiply vector $\mathbf{cf_D}$ to obtain vector **b**:

$$\begin{bmatrix} 0 \\ 3.333333 \\ -2.33333 \end{bmatrix} = \begin{bmatrix} b_A \\ b_B \\ b_C \end{bmatrix} = \begin{bmatrix} -0.001 & 0.001 & 0 \\ 0.0173333 & -0.0006667 & 0 \\ -0.018373 & 0.00070667 & 0.001 \end{bmatrix} \begin{bmatrix} 200 \\ 200 \\ 1200 \end{bmatrix}$$
$$\mathbf{b} = \mathbf{CF}^{-1} \quad \mathbf{cf_d}$$

This simple matrix system yields the same results as our system above. Cash flows, starting with time zero generated by the four-bond arbitrage portfolio, are represented in the following system:

$$\begin{bmatrix} -1000 & -1055.5 & -889 & -1360 \\ 40 & 60 & 0 & 200 \\ 1040 & 60 & 0 & 200 \\ 0 & 1060 & 1000 & 1200 \end{bmatrix} \begin{bmatrix} 0 \\ -3.33333 \\ 2.33333 \\ 1 \end{bmatrix} = \begin{bmatrix} 84 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Thus, by buying Bonds C and D at quantities $2\frac{1}{3}$ and 1, respectively, and shorting Bond B in quantity $3\frac{1}{3}$, the arbitrageur locks in a profit of 84 in time zero, and zero profits in the subsequent three years. The arbitrage portfolio is riskless because all bonds are presumed to be default risk-free and are presumed to be held to maturity, thereby eliminating interest rate and liquidity risk.

8.2 FIXED INCOME HEDGING

Fixed income instruments provide for fixed interest payments at fixed intervals and principal repayments. In the absence of default and liquidity risk (and hybrid or adjustable features), uncertainties in interest rate shifts are the primary source of pricing risk for many fixed income instruments. In this section, we will define a few terms used in the analysis of fixed income instruments and discuss how to manage interest rate uncertainty.

Bond Yields and Sources of Risk

Assume that we wish to analyze a bond maturing in n periods with a *face value* (or principle amount) equal to F paying interest annually at a rate of c. The annual interest payment is rate c multiplied by face value F (or cF) and the bond makes a single payment in

time n equal to F. Using a standard *present value* model discounting cash flows at a rate of k, the bond is evaluated as follows:

$$PV = \sum_{t=1}^{n} \frac{cF}{(1+k)^t} + \frac{F}{(1+k)^n} \tag{8.1}$$

For example, let c equal 0.10, F equal \$1000, k equal 0.12, and n equal 2. The present value of this bond is computed as follows:

$$PV = \frac{100}{(1+0.12)^1} + \frac{100}{(1+0.12)^2} + \frac{1000}{(1+0.12)^2} = 966.20$$

Present value is used to determine the economic worth of a bond; the return of a bond measures the profit relative to the investment of a bond. There are several measures of bond return including current yield and yield to maturity. One such simple measurement, *current yield*, measures the annual interest payments relative to the initial investment required by the bond and is measured as follows:

$$\frac{cF}{P_0} \tag{8.2}$$

If the bond used in our example can be purchased for \$966.20, its current yield will be 10.35%.

An analyst encounters two problems when using current yield as a return measure. First, the current yield does not account for any capital gain or loss ($F - P_0$) that may accrue when the bond matures. Second, current yield does not account for the time value of money or compounding of the cash flows associated with the bond. Hence, the analyst might wish to compute the bond's internal rate of return, which is generally referred to as *yield to maturity* (y):

$$PV = \sum_{t=1}^{n} \frac{cF}{(1+y)^t} + \frac{F}{(1+y)^n} \tag{8.3}$$

Yield to maturity is that value for y that satisfies Equation (8.3) (or k in Equation (8.1)). Usually, a solution must be obtained through an iterative process. The yield to maturity (or internal rate of return) for the bond described above is 12%, computed as follows:

$$P_0 = 966.20 = \frac{100}{(1+0.12)^1} + \frac{100}{(1+0.12)^2} + \frac{1000}{(1+0.12)^2}$$

Thus, yield to maturity can be interpreted as that discount rate which sets the purchase price of a bond equal to its present value.

The *yield to call* for a callable bond is the bond's yield assuming that it will be called or repurchased by its issuer. Yield to call differs from yield to maturity in three respects. First, cash flows are assumed to cease at the call date rather than the maturity date. Second, the call price is used for the yield computation as the bond's final cash flow rather

than the face value of the bond. Third, and only for some quotations, yield to call as calculated based on the call price and call date is quoted only if the bond is selling at a premium (more than face value); otherwise, yield to call is often quoted as yield to maturity instead. The rationale for this more convoluted convention is that if the bond is selling at a discount rather than a premium, it is less likely to be called and more likely to be held to maturity. In fact, *yield to worst* may apply if the bond has one or more call dates, in which case the yield to worst is the lowest of the yield scenarios assuming that the bond is called (or not called) so as to produce the worst possible yield for its owner.

In general, bond risk might be categorized as follows:

1. *Default or credit risk*: The bond issuer may not fulfill all of its obligations.
2. *Liquidity risk*: An efficient market for investors to resell their bonds may not exist.
3. *Interest rate risk*: Market interest rate fluctuations affect values of existing bonds.

In addition, if the bond can be called prior to maturity by its issuer, the bond owner faces call risk.

U.S. Treasury issues are generally regarded as being practically free of default risk. Furthermore, there exists an active market for Treasury issues, particularly those maturing within a short period. Thus, Treasury issues are regarded as having minimal liquidity risk. However, all bonds are subject to interest rate risk. Longer-term bonds are subject to increased interest rate risk due to the increased periods that the yields on longer-term bonds are likely to differ from newly issued bonds. We will discuss hedging this interest risk shortly.

Fixed Income Portfolio Dedication

Fixed income funds seek to ensure relatively stable income levels over given periods of time. Typically, a fixed income fund must provide payments to its creditors, clients, or owners for the given period. For example, a pension fund is often expected to make a series of fixed retirement payments to pension fund participants. Such funds must invest their assets to ensure that their liabilities are paid. In many cases, fixed income funds will purchase assets such that their cash flows exactly match the liability payments that they are required to make. This exact matching strategy is referred to as dedication or laddering and is intended to minimize the risk of the fund. The process of dedication is much the same as the arbitrage transactions discussed above; the fund manager merely determines the cash flows associated with his liability structure and replicates them with a series of default risk-free bonds. For example, assume that a pension fund manager needs to make payments to pension plan participants of $12,000,000 in one year, $14,000,000 in two years, and $15,000,000 in three years. The manager wishes to match these cash flows with a portfolio of Bonds A, B, and C whose characteristics were given above in Table 8.1. These three bonds must be used to match the cash flows associated with the fund's liability structure. For example, in year 1, Bond A will pay $40, B will pay $60, and C will pay $0. These payments must be combined to total $12,000,000. Cash flows must be matched in years 2 and 3 as well. Only one exact matching strategy exists for this scenario. The following system

can be solved for b_A, b_B, and b_C to determine exactly how many of each of the bonds are required to satisfy the fund's cash flow requirements:

$$\begin{aligned} 12,000,000 &= 40b_A + 60b_B + 0b_C \\ 14,000,000 &= 1040b_A + 60b_B + 0b_C \\ 15,000,000 &= 1060b_B + 1100b_C \end{aligned}$$

Note the similarity of this system to the system presented in the previous section. There are many ways to solve this system for the numbers (b) of Bonds A, B, and C to dedicate to the institution's portfolio of liabilities. One way to start is by subtracting the first equation from the second equation, obtaining the following:

$$2,000,000 = 1000b_A$$

This is easily solved for $b_A = 2000$. Now, substitute $b_A = 2000$ into our original second equation:

$$14,000,000 = 1040b_A + 60b_B$$

At this juncture, we see that $b_B = 198{,}666\{2/3\}$. Substitute this result into our original third equation:

$$\begin{aligned} 15,000,000 &= 1060b_B + 1000b_C \\ 15,000,000 &= 210,586,670 + 1000b_C \end{aligned}$$

Now, solve this to obtain $b_C = -195{,}586\{2/3\}$.

Thus, we find from this system that $b_A = 2000$, $b_B = 198{,}666.67$, and $b_C = -195{,}586.67$. We can determine the value of the portfolio of liabilities by weighting their current market prices: $(2000 \times \$1000) + (198{,}666.67 \times \$1055.5) + (-198{,}586.67 \times 889) = \$37{,}816{,}120$. Using matrix notation from Chapter 8 Appendix Section A.2, we have the following system to replicate the portfolio of liabilities:

$$\begin{matrix} \begin{bmatrix} 12,000,000 \\ 14,000,000 \\ 15,000,000 \end{bmatrix} & = & \begin{bmatrix} 40 & 60 & 0 \\ 1040 & 60 & 0 \\ 0 & 1060 & 1000 \end{bmatrix} & \begin{bmatrix} b_A \\ b_B \\ b_C \end{bmatrix} \\ \mathbf{cf_L} & = & \mathbf{CF} & \mathbf{b} \end{matrix}$$

To solve this system, we first invert matrix **CF**, and then use it to premultiply vector $\mathbf{cf_L}$ to obtain vector **b**:

$$\begin{matrix} \begin{bmatrix} 2000 \\ 198666.67 \\ -195586.67 \end{bmatrix} & = & \begin{bmatrix} b_A \\ b_B \\ b_C \end{bmatrix} & = & \begin{bmatrix} -0.001 & 0.001 & 0 \\ 0 \quad .0173333 & -0.0006667 & 0 \\ -0.018373 & 0.00070667 & 0.001 \end{bmatrix} & \begin{bmatrix} 12,000,000 \\ 14,000,000 \\ 15,000,000 \end{bmatrix} \\ \mathbf{b} & = & & \mathbf{CF}^{-1} & & \mathbf{cf_L} \end{matrix}$$

This simple matrix system yields the same results as our system above.

Exact matching or dedication programs can be very effective when liquidity is sufficient to secure appropriately priced bonds with appropriate cash flow structures when needed. However, what if a portfolio manager needs to lock in a cash flow from her portfolio of bonds equal to, say, $1,000,000 on a specific date exactly 12 years, 3 months, and 2 days

into the future? If no bonds are available on that date, some sort of approximation must be acceptable. Alternatively, what if such bonds are available, but overpriced? Again, the manager must work with an approximation or accept investment in overpriced bonds. In any case, dedication programs can limit managers with respect to what bonds they can purchase and in what quantities they must be purchased. Next, we will discuss portfolio immunization, which is not always as effective as an interest-rate hedging tool, but does allow for more flexibility.

8.3 FIXED INCOME PORTFOLIO IMMUNIZATION

As discussed earlier, bonds and certain other debt instruments issued by the U.S. Treasury are often regarded to be practically free of default risk and of relatively low liquidity risk. However, these bonds, particularly those with longer terms to maturity, are subject to market value fluctuations after they are issued, primarily due to changes in interest rates offered on new issues. Generally, interest rate increases on new bond issues decrease values of bonds that are already outstanding; interest rate decreases on new bond issues increase values of bonds that are already outstanding. Immunization models such as the duration model are intended to describe the proportional change in the value of a bond induced by a change in interest rates or yields of new issues.

Bond Duration

Many analysts use present value models to value Treasury issues, frequently using yields to maturity of new Treasury issues to value existing issues with comparable terms. It is important for analysts to know how changes in new-issue interest rates will affect values of bonds with which they are concerned. Bond *duration* measures the proportional sensitivity of a bond to changes in the market rate of interest. Consider a two-year 10% coupon Treasury issue which is currently selling for $966.20. The yield to maturity y of this bond is 12%. Default risk and liquidity risk are assumed to be zero; interest rate risk will be of primary importance. Assume that this bond's yield or discount rate is the same as the market yields of comparable Treasury issues (which might be expected in an efficient market) and assume that bonds of all terms to maturity have the same yield. Further assume that investors have valued the bond such that its market price equals its present value; that is, the discount rate k for the bond equals its yield to maturity y. If market interest rates and yields were to rise for new Treasury issues, then the yield of this bond would rise accordingly. However, since the contractual terms of the bond will not change, its market price must drop to accommodate a yield consistent with the market. Assume that the value of an n-year bond paying interest at a rate of c on face value F is determined by a present value model with the yield y of comparable issues serving as the discount rate k:

$$PV = \sum_{t=1}^{n} \frac{cF}{(1+y)^t} + \frac{F}{(1+y)^n} \tag{8.4}$$

Assume that the terms of the bond contract, n, F, and c are constant. Just what is the proportional change in the price of a bond induced by a proportional change in market interest rates (technically, a proportional change in [1 + y])? This can be approximated by the bond's *Macaulay simple duration formula* as follows:

$$\frac{\Delta PV}{PV} \div \frac{\Delta(1+y)}{(1+y)} \approx Dur = \frac{dPV}{PV} \div \frac{d(1+y)}{(1+y)} = \frac{dPV}{d(1+y)} \times \frac{(1+y)}{PV} \tag{8.5}$$

Equation (8.5) provides a reasonable approximation of the proportional change in the price of a bond induced by an infinitesimal proportional change in (1 + y). To derive this measure of a bond's interest rate sensitivity (Equation (8.5)), we first rewrite Equation (8.4) in polynomial form (to take derivatives later) and substitute y for k (since they are assumed to be equal):

$$PV = \sum_{t=1}^{n} \frac{cF}{(1+y)^t} + \frac{F}{(1+y)^n} = \sum_{t=1}^{n} cF(1+y)^{-t} + F(1+y)^{-n} \tag{8.6}$$

Next, we find the derivative of PV with respect to (1 + y):

$$\frac{dPV}{d(1+y)} = \sum_{t=1}^{n} -tcF(1+y)^{-t-1} - nF(1+y)^{-n-1} \tag{8.7}$$

Equation (8.7) can be rewritten:

$$\frac{dPV}{d(1+y)} = \frac{\sum_{t=1}^{n} -tcF(1+y)^{-t} - nF(1+y)^{-n}}{(1+y)} \tag{8.8}$$

Since the market rate of interest is assumed to equal the bond yield to maturity, the bond's price P_0 will equal its present value PV. Next, multiply both sides of Equation (8.8) by (1 + y) ÷ P_0 to obtain the bond's proportional interest rate sensitivity, which is often more practical for portfolio purposes:

$$Dur = \frac{dPV}{d(1+y)} \times \frac{(1+y)}{P_0} = \frac{\sum_{t=1}^{n} -tcF(1+y)^{-t} - nF(1+y)^{-n}}{P_0} \tag{8.9a}$$

Equation (8.9a) is equivalent to the right side of Equation (8.5). Thus, duration is defined as the proportional price change of a bond induced by an infinitesimal proportional change in (1 + y) or 1 plus the market rate of interest:

$$Dur = \frac{dPV}{d(1+y)} \times \frac{(1+y)}{P_0} = \frac{\sum_{t=1}^{n} \frac{-tcF}{(1+y)^t} + \frac{-nF}{(1+y)^n}}{P_0} \tag{8.9b}$$

Since the market rate of interest will likely determine the yield to maturity of any bond, the duration of the bond described above is determined as follows from Equation (8.9b):

$$Dur = \frac{\frac{-1 \cdot 0.1 \cdot 1000}{(1+0.12)} + \frac{-2 \cdot 0.1 \cdot 1000}{(1+0.12)^2} + \frac{-2 \cdot 1000}{(1+0.12)^2}}{966.20} = -1.907 \tag{8.10}$$

This duration level of −1.907 suggests that the proportional decrease in the value of this bond will equal 1.907 times the proportional increase in market interest rates. This duration level also implies that this bond has exactly the same interest rate sensitivity as a *pure discount bond* (a bond making no coupon payments) that matures in 1.907 years. Duration is always positive. Therefore, an increase in interest rates always decreases the value of a bond. However, although the mathematical value for duration is always negative, and we use that negative value for calculations requiring duration, practitioners normally ignore the minus sign when referring to duration. Its negative value is simply understood, and should never be taken to actually imply a positive relationship between interest rates and bond prices.

Application of the simple Macaulay duration model does require several important assumptions. First, it is assumed that yields are invariant with respect to maturities of bonds; that is, the yield curve is flat. Furthermore, it is assumed that investors' projected reinvestment rates are identical to the bond yields to maturity. Any change in interest rates will be infinitesimal and will also be invariant with respect to time. The accuracy of this model will depend on the extent to which these assumptions hold.

Portfolio Immunization

Earlier, we discussed bond portfolio dedication, which is concerned with matching cash flows of bond portfolios with required payouts associated with liabilities. This process assumes that no transactions will take place within the portfolio and that cash flows associated with liabilities will remain as originally anticipated. Clearly, these assumptions will not hold for many institutions. Alternatively, one may hedge fixed income portfolio risk by using *immunization* strategies, which are concerned with matching the present values of asset portfolios with the present values of cash flows associated with future liabilities. More specifically, immunization strategies are primarily concerned with matching asset durations with liability durations. If asset and liability durations are matched, it is expected that the net fund value (equity or surplus) will not be affected by a very small shift in interest rates; asset and liability changes offset each other. Again, this simple immunization strategy is dependent on the following:

1. Changes in $(1 + y)$ are infinitesimal.
2. The yield curve is flat (yields do not vary over terms to maturity).
3. Yield curve shifts are parallel; that is, short- and long-term interest rates change by the same amount.
4. Only interest rate risk is significant.

The first assumption, because it allows us to use derivatives to measure sensitivities, can only be an approximation when interest rates change by finite amounts. We will discuss bond convexity shortly as a correction for this scenario. Assumptions requiring flat yield curves and parallel yield curve shifts are useful in that we do not have to distinguish between different rates (e.g., short- and long-term rates) over the term of the bond. Immunization becomes significantly more complicated when we need to analyze fixed income risks such as those related to liquidity and default.

Immunization Illustration

Let us revisit our portfolio dedication illustration discussed in the Fixed Income Hedging section. We note here that our illustration allows for a flat yield curve, such that all yields to maturity equal 4%. In this illustration, the pension fund manager still has anticipated cash payouts of $12,000,000, $14,000,000, and $15,000,000 over the next three years 1, 2, and 3. Now, suppose that the manager seeks to immunize interest rate risk associated with this liability stream with a portfolio. Rather than exactly match the liability outflow streams with bond inflows, the manager will match durations of the liability stream with the duration of the bond investment portfolio. This will ensure that changes in the value of the liability stream induced by interest rate changes are approximately the same as changes in the value of the bond portfolio. This will minimize fluctuations in the net value (assets minus liabilities) of the fund as interest rates vary. Bond details were given in Table 8.1. In addition, given the flat yield curve of 4%, the value of the liability stream is $37,817,193.90.

We calculate bond and liability stream durations as follows:

$$Dur_A = \frac{\frac{40}{1+0.04} + 2\times\frac{1040}{(1+0.04)^2}}{-1000} = -1.962$$

$$Dur_B = \frac{\frac{60}{1+0.04} + 2\times\frac{60}{(1+0.04)^2} + 3\times\frac{1060}{(1+0.04)^3}}{-1055.5} = -2.837$$

$$Dur_C = \frac{3\times\frac{1000}{(1+0.04)^3}}{-889} = -3.000$$

$$Dur_L = \frac{\frac{12,000,000}{1+0.04} + 2\times\frac{14,000,000}{(1+0.04)^2} + 3\times\frac{15,000,000}{(1+0.04)^3}}{-37,817,193.90} = -2.048$$

Portfolio immunization is accomplished when the duration (weighted average duration) of the portfolio of bonds equals the duration (−2.048) of the liability stream:

$$\begin{aligned} Dur_A\times w_A + Dur_B\times w_B + Dur_C\times w_C &= Dur_L \\ w_A + \quad w_B + \quad w_C &= 1 \\ -1.962\times w_A - 2.837\times w_B - 3.000\times w_C &= -2.048 \\ w_A + \quad w_B + \quad w_C &= 1 \end{aligned} \tag{8.11}$$

There is an infinity of solutions to this two-equation, three-variable system. Any solution that both satisfies these two equations and satisfies any other of the manager's other constraints and/or preferences is acceptable. Next, suppose that the manager already has invested $3,781,719.39 (10% of the total liability value) into Bond A and wants to hold

this investment in Bond A constant at $w_A = 0.1$. We solve for investment weights as follows:

$$\begin{gathered} -1.962 \times w_A - 2.837 \times w_B - 3 \times w_C = -2.048 \\ w_A + \quad w_B + \quad w_C = 1 \\ w_A = 3,781,719.39/37,817,193.90 = 0.1 \end{gathered}$$

Solving this 3×3 system reveals that the immunized portfolio will have the following weights: $w_A = 0.1$, $w_B = 5.204$, and $w_C = -4.304$. Consider the following system:

$$\begin{bmatrix} -2.048 \\ 1 \\ 0.1 \end{bmatrix} = \begin{bmatrix} -1.962 & -2.837 & -3 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} w_A \\ w_B \\ w_C \end{bmatrix}$$
$$\mathbf{s} = \mathbf{Dur} \quad \mathbf{w}$$

We can solve this system by first inverting matrix **Dur**, then using it to pre-multiply vector **s** to obtain vector **w**:

$$\begin{bmatrix} 0.1 \\ 5.256 \\ -4.356 \end{bmatrix} = \begin{bmatrix} w_A \\ w_B \\ w_C \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 6.135 & 18.405 & -5.840 \\ -6.135 & -17.405 & 4.840 \end{bmatrix} \begin{bmatrix} -2.048 \\ 1 \\ 0.1 \end{bmatrix}$$
$$\mathbf{w} = \mathbf{Dur}^{-1} \quad \mathbf{s}$$

This simple matrix system yields the same results as our system above. Note that small rounding differences can have significant effects on the final portfolio weighting results.

Duration-immunized portfolios are most effective when interest rate changes are infinitesimal. Since interest rate changes are likely to be finite, immunization strategies will be improved if we correct for finite interest rate movements by using convexity. Duration is based on the first derivative of a bond's price with respect to interest rates. This first derivative, or first order approximation, would be accurate only if the relationship were linear, which it is not. To correct for nonlinearities in this relationship, we match asset and liability portfolio convexities as well as durations to correct for finite interest rate changes. We will discuss convexity calculations next.

Convexity

We used duration to determine the approximate change in a bond's value induced by a change in interest rates $(1 + y)$. However, the accuracy of the duration model is reduced by finite changes in interest rates, as we might expect. Duration may be regarded as a first order approximation (it only uses the first derivative) of the change in the value of a bond induced by a change in interest rates. *Convexity* is determined by the second derivative of the bond's value with respect to $(1 + y)$. The first derivative of the bond's price with respect to $(1 + y)$ is given:

$$\frac{\partial P_0}{\partial(1+y)} = \sum_{t=1}^{n} -tcF(1+y)^{-t-1} - nF(1+y)^{-n-1} \tag{8.12}$$

We find the second derivative by determining the derivative of the first derivative as follows:

$$\begin{aligned}\frac{\partial^2 P_0}{\partial(1+y)^2} &= \left[\sum_{t=1}^{n} -t(-t-1)cF(1+y)^{-t-2}\right] - [n(-n-1)F(1+y)^{-n-2}] \\ &= \left[\sum_{t=1}^{n}\frac{(t^2+t)cF}{(1+y)^{t+2}}\right] + \left[\frac{(n^2+n)F}{(1+y)^{n+2}}\right]\end{aligned} \tag{8.13}$$

Convexity is merely the second derivative of P_0 with respect to $(1 + y)$ divided by P_0:

$$Convexity = \frac{\left[\sum_{t=1}^{n}\frac{(t^2+t)cF}{(1+y)^{t+2}}\right] + \left[\frac{(n^2+n)F}{(1+y)^{n+2}}\right]}{P_0} \tag{8.14}$$

The first two derivatives can be used in a second-order Taylor series expansion to approximate new bond prices induced by changes in interest rates as follows:

$$P_1 \approx P_0 + f'(1+y_0)\cdot[\Delta(1+y)] + \frac{1}{2!}\cdot f''(1+y_0)\cdot[\Delta(1+y)]^2 \tag{8.15}$$

$$\begin{aligned}P_1 &\approx P_0 + \left[\sum_{t=1}^{n}\frac{-tcF}{(1+y_0)^{t+1}} - \frac{nF}{(1+y_0)^{n+1}}\right][\Delta y] \\ &\quad + \frac{1}{2}\left[\sum_{t=1}^{n}\frac{(t^2+t)\cdot cF}{(1+y_0)^{t+2}} + \frac{(n^2+n)\cdot F}{(1+y_0)^{n+2}}\right]\cdot[\Delta y]^2 \\ &= P_0 + Dur\cdot\frac{P_0}{1+y_0}\cdot[\Delta y] + \frac{1}{2}\cdot P_0\cdot convexity\cdot[\Delta y]^2\end{aligned} \tag{8.16}$$

Consider a five-year, 10%, $1000-face-value coupon bond currently selling at par (face value). We might compute the present yield to maturity of this bond as $y_0 = 0.10$. The first derivative of the bond's value with respect to $(1 + y)$ at $y_0 = 0.10$ is found from Equation (8.12) to be 3790.78 (duration is $3790.78 \times 1.1 \div 1000 = 4.1698$); the second derivative is found from Equation (8.13) to be 19,368.34 (convexity is $19{,}368.34 \div 1000 = 19.36834$, or rounded at 19.37). If bond yields were to drop from 0.10 to 0.08, the actual value of this bond would increase to 1079.85, as determined from a standard present value model. If we were to use the duration model (first-order approximation from the Taylor expansion, based only on the first derivative), we estimate that the value of the bond increases to 1075.8157. If we use the convexity model second-order approximation from Equation (8.16), we estimate that the value of the bond increases to 1079.6894.

Note that this second estimate with the second-order approximation generates a revised bond value that is significantly closer to the bond's actual value as measured by the present value model. Therefore, the duration and immunization models are substantially improved by the second order approximations of bond prices (the convexity model). The fund manager wishing to hedge portfolio risk should not simply match durations (first derivatives) of assets and liabilities, he should also match their convexities (second derivatives).

Now, let us reconsider our portfolio dedication illustration from Section 8.2 and the portfolio immunization illustration from above. In this illustration, the pension fund manager has anticipated cash payouts of \$12,000,000, \$14,000,000, and \$15,000,000 over the next three years. We calculate bond A, B, and C convexities along with that for the liability stream as follows:

$$Conv_A = \frac{2 \times \frac{40}{(1+0.04)^3} + 6 \times \frac{1040}{(1+0.04)^4}}{1000} = 5.405$$

$$Conv_B = \frac{2 \times \frac{60}{(1+0.04)^3} + 6 \times \frac{60}{(1+0.04)^4} + 12 \times \frac{1060}{(1+0.04)^5}}{-1055.5} = 10.298$$

$$Conv_C = \frac{12 \times \frac{1000}{(1+0.04)^5}}{-889} = 11.095$$

$$Conv_L = \frac{2 \times \frac{12,000,000}{(1+0.04)^3} + 6 \times \frac{14,000,000}{(1+0.04)^4} + 12 \times \frac{15,000,000}{(1+0.04)^5}}{-37,816.120} = 6.375$$

Portfolio immunization is accomplished when the weighted averages of the duration and the convexity of the portfolio of bonds equals the duration and convexity (6.375) of the liability stream:

$$\begin{aligned} Dur_A \times w_A + Dur_B \times w_B + Dur_C \times w_C &= Dur_L \\ Conv_A \times w_A + Conv_B \times w_B + Conv_C \times w_C &= Conv_L \\ w_A + \quad w_B + \quad w_C &= 1 \end{aligned}$$

$$\begin{aligned} -1.962 \times w_A - 2.837 \times w_B - 3 \times w_C &= -2.048 \\ 5.405 \times w_A + 10.298 \times w_B + 11.095 \times w_C &= 6.375 \\ w_A + \quad w_B + \quad w_C &= 1 \end{aligned} \tag{8.17}$$

The single solution to this 3×3 system of equations is $w_A = 0.106$, $w_B = 5.166$, and $w_C = -4.272$. Again, rounding in calculations does affect final results; with less rounding, $w_A = 0.100151$, $w_B = 5.207593$, and $w_C = -4.30774$. Nevertheless, this system provides an improved immunization strategy when interest rate changes are finite.

8.4 TERM STRUCTURE, INTEREST RATE CONTRACTS, AND HEDGING

Thus far in this chapter, we have discussed fixed income arbitrage and hedging without use of derivative instruments such as forward and futures contracts. Also, in Section 8.3, we assumed that yield curves were flat, and that shifts in the yield curve are parallel.

Here, we will discuss long-term interest rates as a function of short-term rates, both originating at time zero (spot rates, based on loans originating now or time zero) and short-term rates originating in the future (forward rates, based on loans originating in the future at rates locked in now). In fact, we will argue that long-term interest rates are related to a combination of short-term rates on loans originating now (spot rates) and on loans originating in the future (forward rates). More specifically, the long-term spot rate will be expressed as a geometric mean of short-term spot and forward interest rates. In addition, we will discuss how to use forward and futures contracts to arbitrage fixed income instruments and to hedge fixed income portfolios. At the end of this section, after having discussed the relationship between long-term spot rates and short-term rates, we will discuss the evolution of short-term rates.

The Term Structure of Interest Rates

The *term structure of interest rates* is concerned with interest rates on debt securities and how these rates vary with respect to varying dates of maturity. For example, term structure might be concerned with why the interest rate on debt maturing in one year is 2% versus 5% for debt maturing in 20 years. Generally at a given point in time, we observe longer-term interest rates exceeding shorter-term rates (*normal yield curves*), although this is not always the case (e.g., the years 1980−1983 when we observed *inverted yield curves*).

The *pure expectations theory* states that long-term spot rates (interest rates on loans originating now) can be explained as a product of short-term spot rates and short-term forward rates (interest rates on loans committed to now but actually originating at later dates). Where $y_{t,t-1}$ is the rate on a one-period, zero-coupon instrument originated at time $t-1$ to be repaid at time t, the pure expectations theory defines the relationship between long- and short-term interest rates as follows:

$$(1+y_{0,n})^n = \prod_{t=1}^{n}(1+y_{t-1,t}) \tag{8.18}$$

Thus, one plus the long-term spot rate $y_{0,n}$ is defined as n-th root minus one of the product of the one-period spot rate $y_{0,1}$ plus one and a series of one-period forward rates $y_{t-1,t}$ plus one. In other words, the long-term spot rate $y_{0,n}$ can be determined based on the short-term spot rate $y_{0,1}$ and a series of one-period forward rates $y_{t-1,t}$ as follows:

$$y_{0,n} = \sqrt[n]{\prod_{t=1}^{n}(1+y_{t-1,t})} - 1 \tag{8.19}$$

Consider an example where the one-year spot rate $y_{0,1}$ is 5%. Investors are expecting that the one-year spot rate one year from now will increase to 6%; thus, the one-year forward rate $y_{1,2}$ on a loan originated in one year is 6%. This means that investors can lock in an interest rate of 6% on a one-year loan originating in one year and maturing in two years. Furthermore, assume that investors are expecting that the one-year spot rate two years from now will increase to 7%; thus, the one-year forward rate $y_{2,3}$ on a loan originated in

two years is 7%. Based on the pure expectations hypothesis, what is the three-year spot rate? This is determined with Equation (8.21) as follows:

$$y_{0,n} = \sqrt[n]{\prod_{t=1}^{n}(1+y_{t-1,t})} - 1 = \sqrt[3]{(1+y_{0,1})(1+y_{1,2})(1+y_{2,3})} - 1$$

$$= \sqrt[3]{(1+0.05)(1+0.06)(1+0.07)} - 1 = 0.05997$$

Term Structure Estimation With Coupon Bonds

The *spot rate* is the yield at present prevailing for zero coupon bonds of a given maturity. The t year spot rate is denoted here by $y_{0,t}$, which represents the interest rate on a loan to be made at time zero and repaid in its entirety at time t. Spot rates can be estimated from bonds with known future cash flows and their current prices. We can obtain spot rates from yields implied from series of coupon bonds if the law of one price holds.

These next few sub-sections are concerned with how interest rates or yields vary with maturities of bonds. The simplest bonds to work with from an arithmetic perspective are *pure discount notes*, that is, notes that make no interest payments. Such notes make only one payment at one point in time—on the maturity date of the note. Determining the relationship between yield and term to maturity for these bonds is quite trivial. The return obtained from a pure discount note is strictly a function of capital gains; that is, the difference between the face value of the note and its purchase price. Short-term U.S. Treasury bills are an example of pure discount (or zero coupon) notes. Coupon bonds are somewhat more difficult to work with from an arithmetic perspective because they make payments to bondholders at a variety of different periods typically every year or every 6 months.

Bootstrapping the Yield Curve

The yield curve can be obtained empirically by examining the payoffs associated with a bond simultaneously with the bond's purchase price. Let D_t be the discount function for time t; that is, $D_t = 1/(1 + y_{0,t})^t$. This means that a coupon payment cF or payment to face value F made at time t will be discounted by multiplying it by the discount function D_t:

$$PV = \sum_{t=1}^{n} cF \times D_t + FD_n = \sum_{t=1}^{n} cF/(1+y_{0,t})^t + F/(1+y_{0,n})^n \tag{8.20}$$

A little algebra on the D_t equality above produces the following spot rate:

$$y_{0,t} = (1/D_t)^{1/t} - 1 \tag{8.21}$$

Thus, one can obtain spot rates $y_{0,t}$ from the bond's current purchase price P_0 and expected future cash flows from coupon payments cF and face value F. Consider a $1000 face value bond making a single interest payment at an annual rate of 5% (see Table 8.2). Suppose this bond is currently selling for 102 (actually meaning 102% of its face value, or

TABLE 8.2 Bootstrapping Spot Rates

Maturity	% Coupon	Ask Price	D_t	Spot Rate
1	5.00	102	0.9714286	2.94%
2	5.00	101-3/4	0.9227891	4.10%
3	5.00	101-1/2	0.8764658	4.49%
4	5.00	101-1/4	0.8323484	4.69%
5	5.00	101-1/4	0.7927128	4.76%
6	5.00	101-1/4	0.7549645	4.80%
7	5.00	101-1/4	0.7190138	4.83%
8	5.00	101-1/4	0.6847751	4.85%
9	5.25	102-1/4	0.6445500	5.00%
10	5.25	102-1/4	0.6123990	5.03%
11	5.25	102-1/4	0.5818518	5.05%
12	5.25	102-1/4	0.5528283	5.06%
13	5.50	104	0.5193962	5.17%
14	5.50	104	0.4923187	5.19%
15	5.50	104	0.4666528	5.21%
16	5.75	105-3/4	0.4331835	5.37%

1020) and that it matures in one year when its coupon payment is made. The one-year spot rate implied by this bond is determined as follows:

$$1020 = (50 + 1000) \times D_1 = (1050)/(1+y_{0,1})^1$$
$$D_1 = 1020/1050 = (0.9714286)^{1/1};\ 1/0.9714286 - 1 = y_{0,1} = 0.0294$$

The one-year spot rate is 2.94%. However, a difficulty arises when the bond has more than one cash flow. As spot rates may vary over time, there may be a spot rate for each period, hence, a spot rate for each cash flow. Now, consider a $1000 face value, two-year, bond making interest payments at an annual rate of 5% (again, see Table 8.2). Suppose that this bond is currently selling for 101.75 (meaning 101.75% or 1017.5) and that it matures in two years when its second coupon payment is made. The two-year spot rate implied by this bond is bootstrapped from the one-year bond as follows:

$$1017.5 = 50 \times 0.9714286 + (50 + 1000) \times D_2$$
$$D_2 = [1017.5 - (50 \times 0.9714286)]/[50 + 1000] = 0.9227891$$
$$(1/.9227891)^{1/2} - 1 = y_{0,2} = 0.0410$$

Bootstrapping simply means to make use of the rate (the one-year rate) or information that is already known to obtain the desired result (the two-year rate). The three-year spot

rate $y_{0,3}$ implied by the three-year bond from Table 8.2 is bootstrapped from the one-year and two-year bonds as follows:

$$1015.0 = 50 \times 0.9714286 + 50 \times 0.9227891 + (50 + 1000) \times D_3$$
$$D_3 = [1015 - (50 \times 0.9714286) - (50 \times 0.9227891)]/[50 + 1000] = 0.8764658$$
$$(1/.8764658)^{1/3} - 1 = y_{0,3} = 0.0449$$

Bootstrapping to map out the n-year yield curve requires that there be one bond maturing in each year t for $t = 1$ to n so that its D_t can be used to determine (bootstrap) the D_t for the bond maturing in one-year subsequent (Figure 8.1). Thus, one starts by determining D_1, then D_2 and so on until all D_t values have been determined through year n.

If we accept the Pure Expectations Theory for the term structure of interest rates, we can obtain forward rates from spot rates. Any i-year forward rate, $y_{t-i,t}$, from year $t - i$ to year t is determined from $(D_t/D_{t-i})^{1/i} - 1$. For example, the two-year spot rate is a function of the one-year spot rate and the one-year forward rate on a loan originating in one year as follows:

$$y_{0,2} = 0.0410 = \sqrt{(1 + y_{0,1})(1 + y_{1,2})} - 1 = \sqrt{(1 + 0.0294)(1 + y_{1,2})} - 1$$

We can use this relationship to solve for the one-year forward rate on a loan originating in one year as follows:

$$(1+0.0410)^2 = (1 + 0.0294)(1 + y_{1,2})$$
$$y_{1,2} = \frac{(1+0.0410)^2}{(1 + 0.0294)} - 1 = 0.052731$$

Similarly, we can solve for the one-year forward rate on a loan originating in two years, forward rate $y_{2,3}$, as follows:

$$(1+0.0449)^3 = (1 + 0.0294)(1 + 0.052731)(1 + y_{2,3})$$
$$y_{2,3} = \frac{(1+0.0449)^3}{(1 + 0.0294)(1 + 0.052731)} - 1 = 0.052744$$

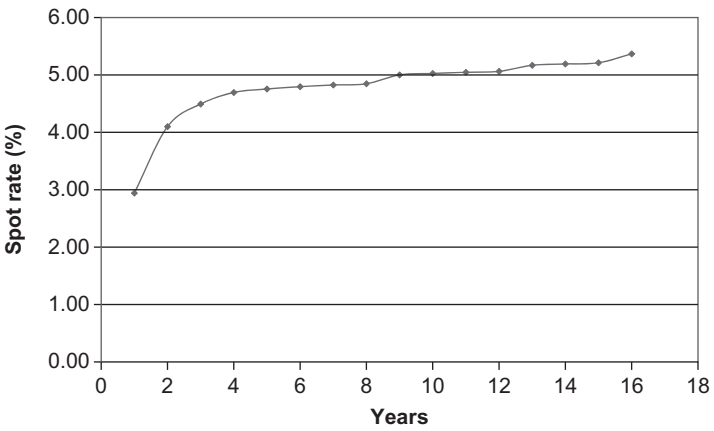


FIGURE 8.1 Mapping the yield curve.

The two-year forward rate on a loan originating in one year, forward rate $y_{1,3}$, is determined as follows:

$$(1+0.0449)^3 = (1+0.0294)(1+y_{1,3})^2$$
$$y_{1,3} = \sqrt{\frac{(1+0.0449)^3}{(1+0.0294)}} - 1 = 0.052737$$

which is identical to

$$y_{1,3} = \sqrt{(1+0.052731)(1+0.052744)} - 1 = 0.052737$$

Simultaneous Estimation of Discount Functions

A coupon bond might be treated as a portfolio of pure discount notes, with each coupon payment being treated as a separate note maturing on the date the coupon is paid. Consider an example involving three bonds whose characteristics are given in Table 8.3. The three bonds are trading at known prices with a total of eight annual coupon payments among them (three for bonds F and G and two for bond E). Bond yields or spot rates must be determined simultaneously to avoid associating contradictory rates for the annual coupons on each of the three bills.

Let D_t be the discount function for time t; that is, $D_t = 1/(1 + y_{0,t})^t$. Since $y_{0,t}$ is the spot rate or discount rate that equates the present value of a bond with its current price, the following equations may be solved for discount functions then spot rates:

$$947.376 = 50D_1 + 1050D_2$$
$$904.438 = 60D_1 + 60D_2 + 1060D_3$$
$$981 = 90D_1 + 90D_2 + 1090D_3$$

This system of equations might be represented by the following system of matrices:

$$\underset{\mathbf{CF}}{\begin{bmatrix} 50 & 1050 & 0 \\ 60 & 60 & 1060 \\ 90 & 90 & 1090 \end{bmatrix}} \quad \underset{\mathbf{d}}{\begin{bmatrix} D_1 \\ D_2 \\ D_3 \end{bmatrix}} = \underset{\mathbf{P}_0}{\begin{bmatrix} 947.376 \\ 904.438 \\ 981 \end{bmatrix}}$$

To solve this system we first invert matrix CF, and then use this inverse to premultiply vector P_0 to obtain vector d:

$$\underset{\mathbf{CF}^{-1}}{\begin{bmatrix} -0.001 & -0.03815 & 0.0371 \\ 0.001 & 0.00181667 & -0.00177 \\ 0 & 0.003 & -0.002 \end{bmatrix}} \quad \underset{\mathbf{P}_0}{\begin{bmatrix} 947.376 \\ 904.438 \\ 981 \end{bmatrix}} = \underset{\mathbf{d}}{\begin{bmatrix} D_1 \\ D_2 \\ D_3 \end{bmatrix}} = \begin{bmatrix} 0.943377 \\ 0.85734 \\ 0.751316 \end{bmatrix}$$

TABLE 8.3 Coupon Bonds E, F, and G

Bond	Current Price	Face Value	Coupon Rate	Years to Maturity
E	947.376	1000	0.05	2
F	904.438	1000	0.06	3
G	981	1000	0.09	3

Thus, we find from solving this system for vector d that $D_1 = 0.943377$, $D_2 = 0.85734$, and $D_3 = 0.751316$. Since $D_t = 1/(1 + y_{0,t})^t$, $1/D_t = (1 + y_{0,t})^t$, and $y_{0,t} = 1/D^{1/t} - 1$. Thus, spot rates are determined as follows:

$$\frac{1}{D_1} - 1 = 0.06 = y_{0,1}$$

$$\frac{1}{D_2^{1/2}} - 1 = 0.08 = y_{0,2}$$

$$\frac{1}{D_3^{1/3}} - 1 = 0.10 = y_{0,3}$$

Note that there exists a different spot rate (or discount rate) for each term to maturity; however, the spot rates for all cash flows generated by all bonds at a given period in time are the same. Thus, $y_{0,t}$ will vary over terms to maturity, but will be consistent for all bonds in a given time period. Thus, for example, the values of the three bonds are determined as follows:

$$947.376 = 50D_1 + 1050D_2 = \frac{50}{(1+0.06)^1} + \frac{1050}{(1+0.08)^2}$$

$$904.438 = 60D_1 + 60D_2 + 1060D_3 = \frac{60}{(1+0.06)^1} + \frac{60}{(1+0.08)^2}\frac{1060}{(1+0.10)^3}$$

$$981 = 90D_1 + 90D_2 + 1090D_3 = \frac{90}{(1+0.06)^1} + \frac{90}{(1+0.08)^2}\frac{1090}{(1+0.10)^3}$$

Any other bond with cash flows paid at the ends of some combination of years one, two, and three must have a value that is consistent with these three spot rates. For example, a three-year 2% coupon Bond H must be valued as follows:

$$P_H = 20D_1 + 20D_2 + 1020D_3 = \frac{20}{(1+0.06)^1} + \frac{20}{(1+0.08)^2}\frac{1020}{(1+0.10)^3} = 802.36$$

A different market price for Bond H will lead to an arbitrage opportunity. In addition, prices for Bonds E, F, and G above will lead to the following forward rates:

$$y_{1,2} = \frac{(1+0.08)^2}{(1+0.06)} - 1 = 0.10$$

$$y_{2,3} = \frac{(1+0.10)^3}{(1+0.06)(1+0.10)} - 1 = 0.14$$

$$y_{1,3} = \sqrt{\frac{(1+0.10)^3}{(1+0.06)}} - 1 = 0.12$$

Again, contract or bond contracts that are inconsistent with these rates lead to arbitrage opportunities.

The Evolution of Short-Term Rates[2]

As we discussed earlier in this section, the yield curve depicts varying spot rates over associated terms to maturity. Understanding the nature of the uncertainty that drives spot and forward rates, particularly short term rates is essential to understanding fixed income instruments. The Merton (1973) term structure model prices bonds based on the assumption that short-term interest rates (more precisely, instantaneous spot r_0 and forward rates r_t) are related to an arithmetic Brownian motion process Z_t, similar to the stock price model in Section 7.1:

$$dr_t = \mu dt + \sigma dZ_t$$

Instantaneous rates r_t following arithmetic Brownian motion are normally distributed, and normal distributions are often very easy to work with. On the other hand, changes in interest rates are unrelated to historical or long-term mean rates $\bar{r}$. This means that directional moves for short term interest rates cannot be predicted based on available information, particularly when the drift μ is low compared to interest rate volatility σ. The range of potential interest rate changes is from an unreasonable negative infinity to positive infinity. Negative interest rates are very possible under Brownian motion, but may seem less likely (though not impossible) in practice.

These problems that can produce negative interest rates and constant interest rate volatility can be alleviated with the Rendleman and Bartter (1980) geometric Brownian motion process model for short-term interest rates:

$$dr_t = \mu r_t dt + \sigma r_t dZ_t$$

In this model, the long-term interest rate is lognormally distributed with drift parameter μ and volatility parameter σ, a handy quality for modeling in continuous time. As interest rates r_t approach zero, the drift μr_t and rate volatility σr_t approach zero, and rates will not turn negative. Volatility and drift will increase as rates increase from zero in this geometric Brownian motion process.

Unfortunately, neither of these models capture the tendency for short term interest rates to revert towards some long-term mean rate $\bar{r}$ as we often observe in actual financial markets. This long-term mean interest rate $\bar{r}$ is not necessarily (though could be) the same as the long-term spot rate that we discussed earlier, but the mean of short-term rates over an extended period of time. For example, we might observe that when the short-term spot rate exceeds the long-term mean rate ($r_0 > \bar{r}$), the drift might be expected to be negative so that the short-term rate drifts down towards the long-term mean rate. We might say that the interest rates are currently high in this scenario, and we expect for them to drop towards the long-term mean. When the short-term rate r_0 is less than the long-term rate ($r_0 < \bar{r}$), the drift might be expected to be positive. Thus, the short-term rate has a tendency to revert to its long-term mean $\bar{r}$, whose value might be the value justified by economic fundamentals such as capital productivity, long-term monetary policy, etc.

[2]This subsection, intended as a very superficial introduction to term structure models can be skipped without loss of continuity in the chapter.

Define the term $0 < \lambda < 1$ to be a constant that reflects the speed of the mean-reverting adjustment for the instantaneous rate r_t towards its constant long-term mean rate $\overline{r}$; that is, λ is the mean reversion factor, sometimes called a "pullback factor." This pullback factor is typically estimated or calibrated based on a statistical analysis of historical data. Let σdZ_t represent random shocks or disturbances to r_t. If volatility σ is assumed to be independent of the short-term rate (e.g., it is a constant), the following defines the *Ornstein-Uhlenbeck* (Uhlenbeck & Ornstein, 1930) mean reverting process, also known as a *Vasicek process* (Vasicek, 1977) in fixed income analysis:

$$dr_t = \lambda(\overline{r} - r_t)dt + \sigma dZ_t$$

The Ornstein-Uhlenbeck process is sometimes called an elastic random walk. The Ornstein-Uhlenbeck process has two components, the mean reversion component $\lambda(\overline{r} - r_t)$ and the Brownian motion component σdZ_t. The Brownian motion component is the disturbance factor that causes the short-rate r_t to diverge from the long-term mean rate $\overline{r}$. The mean reversion component draws the short term rate r_t back towards the long term mean rate $\overline{r}$. A higher value for λ implies a faster reversion ($\lambda < 1$) by the short-term rate r_t to the long term mean rate $\overline{r}$.

If λ were zero, there would be no mean reversion and the process would be a Brownian motion. One drawback to the Vasicek interest rate model is that interest rate shifts have a normally distributed component, leading to the unfortunate result that it is possible for the interest rate to become negative. Obviously, this creates an arbitrage opportunity when cash is available for investors to hold. Figure 8.2 depicts a simulation of a Vasicek process over length of time 200, with $r_0 = 0.05$, and $\sigma = 0.02$ and $\lambda = 0.1$. Also notice in Figure 8.2 that the process can drop below zero.

The Vasicek yield curve model has a number of desirable characteristics. The model captures the empirical tendency for interest rates to revert towards some sort of mean rate. The model is driven off short term interest rates, much as actual interest rates might be impacted by the Federal Funds rate, the "overnight" bank-to-bank controlled by the central bank (Fed). However, there are a number of problems with the Vasicek model in characterizing the behavior of the yield curve:

1. The Vasicek model is likely to apply only in reasonably "normal" scenarios. For example, in situations involving crises such as hyperinflation, mean reversion is not likely to characterize the behavior of interest rates.
2. Because it is based on Brownian motion, the Vasicek model does not allow for discrete jumps in the interest rate process.
3. The Vasicek model produces the result that all short- and long-term rate shifts over terms to maturity are perfectly correlated.

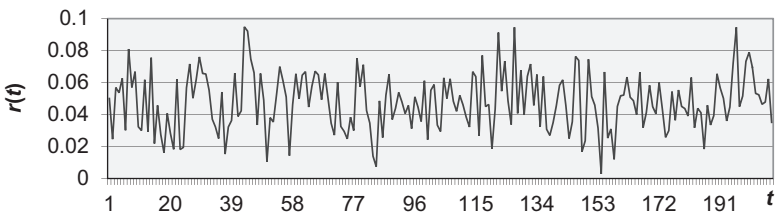


FIGURE 8.2 Simulation of Vasicek process: $r_0 = 0.05$, $\sigma = 0.02$ and $\lambda = 0.1$.

4. Related to the difficulty put forth just above, the Vasicek model assumes only a single underlying risk factor when, in fact, there is significant evidence that there may well be multiple factors. For example, sometimes the yield curve can "twist;" that is, long- and short-term rates can move in opposite directions. Multiple risk factors can often explain such "twisting."
5. Finally, the Vasicek model allows for the possibility of negative interest rates, even for negative real interest rates, a phenomenon that we should expect to observe rarely, if at all. We discussed a solution to this problem earlier (Bartter and Rendleman), and will introduce the Cox-Ingersoll-Ross model shortly as an alternative.

Why work with an interest rate model that presents all of these difficulties? As with most other financial models, we simply balance realism and ease of model building. The Vasicek model does capture some of the characteristics of a reasonable interest rate process and it is rather easy to work with, particularly in terms of parameter calibration. In addition, it is useful and sometimes very straightforward to adapt this framework into more realistic alternative depictions of interest rate processes.

The "square root correction" in the Cox Ingersoll Ross Model (CIR: Cox, Ingersoll, & Ross, 1985) eliminates the potential problem of negative interest rates in the mean reverting Vasicek model:

$$dr_t = \lambda(\overline{r} - r_t)dt + \sigma_r\sqrt{r_t}dZ_t.$$

As in the Vasicek model, the mean reversion component draws the short-term rate back towards the long-term rate. However, notice that the nominal (not proportional) volatility of rates is not constant; it declines as the interest rate declines. That is, as the short term rate r_t approaches zero, the random component of interest rate changes $\sigma\sqrt{r_t}dZ_t$ will also approach zero. First, this characteristic is more realistic for interest rates, as the volatility of rates tends to be reduced as interest rates approach zero. Second, this process ensures that rates will not drop to or below zero, eliminating arbitrage opportunities for investors holding cash. Both represent advantages of CIR over the Vasicek model.

8.5 ARBITRAGE WITH CURRENCIES

Currency arbitrage can exist when currency prices are not consistent with respect to other currencies or with respect to goods or financial instruments in different countries. Thus, there exist many potential types of currency (foreign exchange or FX) arbitrage opportunities. Here, we discuss triangular arbitrage and absolute purchase power parity (PPP), scenarios that often do not require derivative securities to exploit. When no-arbitrage conditions force prices to retain consistency, triangular arbitrage and PPP can be used to hedge currency risk.

Triangular Arbitrage

Triangular arbitrage exploits the relative price difference between one currency and two other currencies. Suppose, for example, the buying and selling prices of EUR1.0 is USD1.2816. However, another currency, the South African Rand (SFR), has a price (buying or selling) equal to USD0.2000 or EUR0.1600. Is there a triangular arbitrage opportunity?

There is if there is a price discrepancy for SFR in terms of USD and EUR. Note that the price of SFR in terms of both dollars and euro imply that 1 EUR is worth 1.25 USD:

$$\text{SFR1} = \text{USD0.2000} = \text{EUR0.1600}$$

Since USD0.20 = EUR0.16, dividing both figures by 0.16 implies that USD1.25 = EUR1. But, this is inconsistent with the currency price information given above, which states that USD1.2816 = EUR1.0. Even without knowing which pair of currencies is relatively mispriced, we can exploit an arbitrage opportunity taking advantage of the price inconsistencies. So, we will start rather arbitrarily with one pair, the SFR and USD. In terms of the SFR, it appears that the USD is too strong relative to the EUR, so we will start by selling USD0.20 for SFR1 as per the price given above. We will cover the short position in USD by selling EUR0.16, which actually nets us $0.16 \times \text{USD1.2816} = \text{USD0.2051}$. We will cover our short position in EUR by selling SFR at the price listed above. Positions from our three transactions are summarized and totaled in Table 8.4.

Because of our price inconsistency, we will earn USD0.0051 every time that we execute this series of transactions. Existence of such arbitrage profits should be exploited by arbitrageurs, whose arbitrage transactions will ultimately force down the price SFR in terms of USD and force up the price of SFR in terms of EUR until the discrepancy is eliminated.

Purchase Power Parity, Arbitrage, and Hedging in FX Markets

In a world economy characterized by free trade, complete certainty and no market frictions (such as transactions costs), we would expect that exchange rates will be regulated by *purchase power parity* (PPP). PPP is an example of the *law of one price*, which simply states that securities offering the same cash flow characteristics or baskets offering the same commodities must sell for the same price. PPP states that the values of currencies relative to each other are determined by the quantity of goods they will purchase. This condition holds both in spot markets and forward markets (we will discuss forward markets in the next section). Currency traders creating arbitrage portfolios should force the PPP to hold. Violations of this parity condition should lead traders to purchase the less expensive currency sell the more expensive currency.

One well-known (though somewhat tongue in cheek) test of PPP is the "Big Mac standard" popularized by *The Economist*. MacDonald's Corporation's Big Mac hamburgers are generally regarded to be more or less identical all over the world. Given PPP, then the Big Mac should sell for the same price in each country. For example, suppose that the Big Mac cost $4.60 in a U.S. restaurant and costs £3.62 (GBP3.62) in the UK; Further suppose that

TABLE 8.4 Triangular Arbitrage

	USD	SFR	EUR
Sell USD0.20 for SFR1	− 0.2000	+ 1.0000	
Sell EUR0.16 for USD0.2051	+ 0.2051		− 0.1600
Sell SFR1.0 for EUR0.16		− 1.0000	+ 0.1600
Totals	+ 0.0051	0	0

the dollar/pound exchange rate is \$1.46/£; that is, GBP1 has the same value as USD1.46. At this exchange rate of \$1.46/£, the British pound appeared overvalued by approximately 15% relative to the U.S. dollar. However, we need to recognize that Big Macs are not easily exported from countries where they are underpriced, which might prevent absolute PPP from holding in this instance. Nor does this relationship among prices account for differences in taxes, subsidies, labor, and other production costs.

Purchase Power Parity in Spot Markets

Purchase power parity states that two commodities must sell for the same price (after adjusting for relative currency values). PPP is the primary basis for international comparisons of prices, income and expenditures. In spot markets, the price of a commodity in a foreign market (e.g., P_{UK0}—the price in the UK) multiplied by the exchange rate at present (S_0—the dollars per pound exchange rate) should equal the price of the commodity in domestic markets (e.g., P_{US0}—the price in the United States):

$$P_{US0} = S_0 \times P_{UK0}$$

This condition is known as PPP in spot markets or absolute PPP. In the absence of market frictions in both commodity and exchange markets, and with the ability to freely transport commodities between markets, the potential for arbitrage should force absolute PPP in spot markets to hold. For example, if the spot price of gold in U.S. markets is USD1400 per ounce and UK1000 in British markets, the spot exchange rate must be USD1.400/GBP (Great Britain Pounds). Any single deviation from these rates will lead to an arbitrage opportunity. We will discuss changes in exchange rates and forward markets for exchange in the next section.

8.6 ARBITRAGE AND HEDGING WITH CURRENCY FORWARD CONTRACTS

In frictionless markets, currency arbitrage opportunities exist when forward currency exchange rates are inconsistent given prices of commodities or rates associated with financial instruments in different countries. Here, we will discuss triangular arbitrage in forward markets and relative PPP, scenarios that do require certain derivative securities.

Parity and Arbitrage in FX Markets

In a world economy characterized by free trade, complete certainty, and no market frictions (such as transactions costs), we can argue that the following conditions will hold in equilibrium:

1. Purchase power parity (PPP)
2. Interest rate parity (IRP)
3. Forward rates equal expected spot rates
4. The Fisher effect
5. The international Fisher effect

Collectively, these conditions are often referred to as the *international equilibrium model*. Violations of these conditions imply either arbitrage opportunities or frictions that prevent the exploitation of arbitrage opportunities.

Purchase Power Parity

We argued above that PPP states that two commodities must sell for the same price (after adjusting for relative currency values). In forward markets, as in spot markets, the price of a commodity in a foreign market (e.g., P_{UK1}—the forward price in the UK) multiplied by the forward exchange rate (F_1—the dollars per pound forward exchange rate) should equal the price of the commodity in domestic markets (e.g., P_{US1}—the forward price in the United States):

$$P_{US1} = F_1 \times P_{UK1}$$

This condition is known as PPP in forward markets. In the absence of market frictions in both commodity and exchange markets, and with the ability to freely transport commodities between markets, the potential for arbitrage should force absolute PPP in spot or forward markets to hold.

The combination of PPP conditions in both spot and forward markets leads to the *relative PPP* condition. Consider a basket of groceries in the United States, which costs 20 dollars ($P_{US0} = \$20.00$). Assume that the same basket of groceries can be purchased in the United Kingdom for 12.50 pounds ($P_{UK0} = £12.50$). Further assume that the exchange rate of dollars for British pounds (S_0) is 1.6; that is, 1.6 dollars buys one pound. Purchase power parity holds here in the spot market, since an American can exchange 20 dollars for 12.5 pounds in the foreign exchange market and purchase the same basket of commodities as he could in the United States.

If in one year, the price of the basket were to increase by 10% to USD22 (perhaps due to inflation) in the United States, and increase by 8% in the United Kingdom, what would we expect to see happen to the exchange rate of dollars for pounds (F_1)? Recall that the PPP Theorem states that the same basket of commodities must cost the same in any two economies, given relevant exchange rates. Furthermore, changes in price levels in one economy will be matched by either offsetting changes in its currency value (exchange rates), changes in prices in the second economy, or some combination of the two. Thus, since the price of goods in the United States increased by 10% and the price of goods in the United Kingdom increased by 8%, the exchange rate should increase by 1.8512% to 1.6296 dollars for one pound ($F_1 = 1.6296$). Thus, the change in the exchange rate (ΔS) under PPP must be:

$$\Delta S = \frac{F_1}{S_0} - 1 = \frac{1.6296}{1.6000} - 1 = \frac{P_{US1}/P_{US0}}{P_{UK1}/P_{UK0}} - 1 = \frac{22.00/20.00}{13.50/12.50} - 1 = 0.01851. \tag{8.22}$$

Note that the country with the higher inflation rate will have its currency devalued relative to the country with the lower inflation rate. More generally, if (π_{US}) represents the U.S. inflation rate and (π_{UK}) represents the UK inflation rate, then the following holds:

$$\frac{(1 + \pi_{US})}{(1 + \pi_{UK})} = \frac{F_1}{S_0} \tag{8.23}$$

Equation 8.23 is the general mathematical definition or statement of PPP. It provides the relationship between forward and spot rates of exchange as a function of relative inflation rates in relevant countries.

Example 1: Purchase Power Parity Violation and Arbitrage

Consider the following violation of PPP. The exchange rates of dollars for pounds are 1.6000 and 1.6296 in the spot and one-year forward markets, respectively. Assume that gold is selling for $400 per ounce in American spot markets and for £250 in British spot markets. The one-year forward price of gold is $440 in American markets and £280 in British markets. We might wish to assume that inflation rates in the United States and the UK are, respectively, 10% and 12%. In this case, we should be able to demonstrate an arbitrage opportunity. Consider the following transactions in forward markets (future prices are locked in now by taking positions in forward contracts):

Transaction Number	Transaction
1	Long forward contract for one ounce of gold in United States
2	Short forward contract for £270 at $F = 1.6296$
3	Short forward contract for one ounce of gold in UK

Transaction Number	Time-One Gold Position	Time-One Pound Position	Time-One Dollar Position
1	+1 ounce		−$440
2		−£270	+$440
3	−1 ounce	+£280	
Totals	0	+£10	

Because all prices are locked in by spot and forward contracts, we are able to lock in a profit of £10 by engaging the above transactions. The change in exchange rates does not coincide appropriately with the countries' relative gold price changes. You should be able to demonstrate for yourself that Equation (8.23) does not hold for this example. Whenever Equation (8.23) does not hold, we should either be able to exploit an arbitrage opportunity or identify a constraining market friction.

How do we know that we should purchase gold in the United States at time one? Consider the following:

$$\frac{(1+\pi_{US})}{(1+\pi_{UK})} = \frac{(1+0.10)}{(1+0.12)} < \frac{F_1}{S_0} = \frac{1.6296}{1.6} \tag{8.24}$$

The forward price of the pound ($1.6296) increased relative to the dollar while the inflation rate affecting the dollar is less than for the pound. This means that the forward price of gold in the United States is too low relative to the UK, the forward price of gold in the UK is too high relative to the United States, and the forward price of pounds is too high relative to dollars.

Interest Rate Parity

Interest rate parity states that anticipated currency exchange rate shifts will be proportional to countries' relative interest rates. Continuing the above example, assume that the current nominal interest rate in the United States is 12%, and the spot exchange rate of dollars for pounds is 1.6. Based on these rates, what must be the nominal interest rate in the UK? To avoid arbitrage opportunities in the market for interest bearing securities, interest rate parity must hold:

$$\frac{(1+i_{US})}{(1+i_{UK})} = \frac{F_1}{S_0} = \frac{(1+0.12)}{(1+i_{UK})} = \frac{1.6296}{1.600} \tag{8.25}$$

Thus, the interest rate in the UK must be 9.9636%. If the British interest rate were lower, arbitragers would borrow at the lower British rate, exchange pounds for dollars, and then loan at the higher American interest rate. Thus, interest rate parity holds that a strategy of borrowing money in one currency, immediately exchanging that currency for a second that is immediately loaned, will enable the investor to break even on the loans if futures contracts are used to lock in future exchange rates.

Example 2: Interest Rate Parity Violation and Arbitrage

Here, we continue to assume that exchange rates of dollars for pounds are 1.6 and 1.6296 in the spot and one-year forward markets, respectively. Assume that nominal interest rates are 12.5% in the United States and 12% in the United Kingdom. Again, we should be able to demonstrate an arbitrage opportunity. Consider the following transactions in markets for interest bearing securities and forward exchange (forward prices are locked in now by contracts):

Transaction Number	Transaction
1.	Borrow $1000 now in the United States at 12.5%; repay at Time One
2.	Buy £625 now for $1000
3.	Loan £625 at 12%; Collect proceeds at Time One
4.	Sell £700 at Time One at $F_1 = 1.629629$ for $1140.74

Time Zero Positions

Transaction Number	Time Zero Pound Position	Time Zero Dollar Position
1		+$1000
2	+£625	−$1000
3	−£625	
4	_____	_____
Totals	0	0

Time One Positions

Transaction Number	Time One Pound Position	Time One Dollar Position
1		−$1125.00
2		
3	+£700	
4	− £700	+ $1140.74
Totals	0	+15.74

Here, we are able to lock in a profit of $15.74 by engaging the above transactions. The change in exchange rates did not coincide appropriately with the countries' relative interest rates. You should be able to demonstrate for yourself that Equation (8.24) does not hold for this example. Whenever Equation (8.24) does not hold, we will be able to locate an arbitrage opportunity or a friction that prevents its exploitation.

Exchange Rate Expectations

The third of the four parity relations states that the prevailing forward rate of exchange equals the expected spot rate at the time of delivery:

$$F_1 = E[S_1] \tag{8.26}$$

Speculators will force Equation (8.26) to hold, at least approximately, through their trading activity. However, its violation does not necessarily represent an arbitrage opportunity.

The Fisher Effect

The *Fisher effect* within a given country states that the relationships among the nominal interest rate (i), the real interest rate (i') and the expected inflation rate $E(\pi)$ are given by Equation (8.27):

$$(1 + i) = (1 + i')(1 + E(\pi)) \tag{8.27}$$

The Fisher effect provides a definition for the real rate i' of interest in an economy in terms of the nominal rate i and the inflation rate π. Equation (8.27) provides a definition, but does not necessarily indicate an arbitrage opportunity.

The International Fisher Effect

The international counterpart to the Fisher effect is the final parity condition—the *international Fisher effect*. The international Fisher effect draws from the Fisher effect and purchase and interest rate parity. For example, the following states the international fisher effect for the United States, the United Kingdom:

$$\frac{(1 + i_{US})}{(1 + i_{UK})} = \frac{(1 + E[\pi_{US}])}{(1 + E[\pi_{UK}])} \times \frac{(1 + i'_{US})}{(1 + i'_{UK})} \tag{8.28}$$

Thus, the Fisher effect specifies the relationships among real and nominal interest rates and inflation rates. The international Fisher effect extends this relation among countries, indicating that real interest rates among countries will be identical. If this condition were not to hold, arbitrageurs would borrow in countries with low real interest rates and lend in countries with higher nominal interest rates. Alternatively, they might be able to exploit arbitrage opportunities arising from violations in PPP. Please take time to notice the relationships among Equation (8.23) through (8.28) and how additional equalities can be created from them.

8.7 HEDGING EXCHANGE EXPOSURE

Multinational companies face a number of risks not experienced by companies operating in only one country. Among the most significant of these risks is exchange rate uncertainty and fluctuations. Transaction exposure is a firm's risk arising from settlement of obligations denominated in foreign currencies. The following are sources of transaction exposure:

1. Purchasing or selling goods or services on foreign-denominated credit
2. Borrowing or lending with foreign-denominated notes
3. Taking positions in forward exchange contracts
4. Buying or selling assets denominated in foreign currencies

The corporate treasurer is responsible for managing these risks. The treasurer will either employ traders or engage agency traders to take exchange and derivative positions to manage these risks. The following are strategies that treasurers and traders employ to contend with transactions exposure:

1. Do nothing—accept the risk
2. Hedge in forward markets
3. Hedge in money markets: take offsetting position in local currency debt instruments
4. Hedge in option markets: incomplete hedge with puts or complete hedge with conversions

We use illustrations to examine the first three of these strategies for dealing with transactions exposure. We will discuss the options strategies in the next chapter.

Illustration: Managing Transactions Exposure

Consider an example regarding the management of transactions exposure. The Dayton Company of America invested in a British construction project and expects to receive a payoff of £1,000,000 in three months. The company seeks to realize its revenues in dollars. The problem is that management does not know the dollar amount of the payoff due to exchange rate uncertainty. Assume relevant data as follows:

Spot exchange rate: $1.7640/£
Three-month forward exchange rate: $1.7540/£
UK borrowing interest rate: 10.0%

U.S. borrowing interest rate: 8.0%
UK lending interest rate: 8.0%
U.S. lending interest rate: 6.0%
Forward contract settlement price: $1.75 per £
Transactions cost on £1,000,000 forward contract: $500

The treasurer's problem is to evaluate non-options–based methods of managing the transaction risk associated with this extension of credit and the implications of each. We will determine which hedging strategy is likely to be optimal and why. First, we consider the alternative of doing nothing.

Unhedged Alternative

The unhedged alternative simply means to accept the risk. The strategy is simply to wait 3 months, then sell £1,000,000 for dollars at the then prevailing spot rate. The result of this strategy is that all £1,000,000 is at risk. The expected value of the transaction is $1,754,000 based on the forward rate of exchange used as a predictor for the future spot rate. However, all funds are at risk.

Forward Market Hedge

The *forward market hedge* strategy is to sell £1,000,000 forward for dollars at once. The result is that, in the absence of default risk, $1,754,000 will certainly be received in three months. Transactions costs at time zero will total $500. Forgone interest on this $500 over three months totals $\$7.50 = 0.06 \times 0.25 \times \$500$. The total amount (net of forgone interest) to be received in three months is $1,753,492.50. This amount is certain.

Money Market hedge

The *money market hedge* strategy is to borrow £975,609.76 in UK for three months at an annual rate of 10%, exchange the £975,609.76 for $1,720,975.60 now at the spot exchange rate, and then invest $1,720,975.60 for three months at an annual rate of 6%. The result of this strategy is that the UK pound loan is repaid by receipts from sale in three months. $1,746,790.20 is obtained from the U.S. investment. This amount is certain in the absence of default risk. Note that this strategy in this example is inferior to the forward market hedge.

Other Hedging Strategies

The hedging strategies described in the above examples and in the next chapter are more effective when firms can easily enter into forward, money market and options contracts. This is normally quite easily accomplished in countries with major currencies. However, in many instances, firms will need to hedge exchange rate volatility in countries where such opportunities are not available. For example, it is often very difficult to employ the above strategies in smaller African, Asian, or Latin American countries. In many instances, firms might modify these strategies with cross-hedges. Typically, these cross-hedging strategies involve the use of contracts denominated in currencies strongly correlated with the currency to be hedged. For example, rather than attempt to directly hedge exchange rate risk involving Philippine pesos, a firm might use contracts denominated in

yen, which are fairly highly correlated with pesos. An alternative cross hedging strategy is to hedge currency risk with commodity contracts whose values are strongly correlated with that of the currency. For example, the price of oil is strongly correlated with the value of the Saudi Arabia riyal (SAR). The firm needing to hedge risk associated with the riyal can use futures contracts on oil as an imperfect substitute for the currency itself.

Additional Reading

Elton, Gruber, Brown, and Goetzmann (2010), Chapters 21 and 22, offer good coverage of yield curves, fixed income arbitrage, and hedging. Knopf and Teall (2015) offer a somewhat more detailed introduction to much of the fixed-income-related material in this chapter, as does Chapter 18 in Neftci (2000), which extends material in this chapter to continuous time. Chapters 1 through 6 in Neftci comprise good background reading for Chapter 18. Shapiro (2006) and Grabbe (1996) discuss parity conditions along with currency exposure and hedging. Chapter 9 in Saunders and Cornett (2009) provides a concise overview of FX markets, instruments and analysis.

References

Cox, J., Ingersoll, J., & Ross, S. (1985). A theory of the term structure of interest rates. *Econometrica, 53*, 385–408.

Elton, E., Gruber, M., Brown, S., & Goetzmann, W. (2010). *Modern portfolio theory and investment analysis* (8th ed.). New York: Wiley.

Grabbe, J. O. (1996). *International financial markets* (3rd ed). Englewood Cliffs, NJ: Prentice Hall.

Knopf, P., & Teall, J. L. (2015). *Risk neutral pricing and financial mathematics: A primer*. San Diego: Academic Press.

Merton, R. C. (1973). Theory of rational option pricing. *Bell Journal of Economics and Management Science, 4*(1), 141–183.

Neftci, S. N. (2000). *An introduction to the mathematics of financial derivatives* (2nd ed). San Diego: Academic Press.

Rendleman, R. J., Jr., & Bartter, B. J. (1980). The pricing of options on debt securities. *Journal of Financial and Quantitative Analysis, 15*(1), 11–24.

Saunders, A., & Cornett, M. M. (2009). *Financial markets and institutions: A modern perspective*. Irwin, Boston: McGraw-Hill.

Shapiro, A. C. (2006). *Multinational financial management* (8th ed). New York: Wiley.

Uhlenbeck, G. E., & Ornstein, L. S. (1930). On the theory of Brownian motion. *Physics Review, 36*, 823–841.

Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics, 5*, 177–188.

8.8 EXERCISES

1. There are two three-year bonds with face values equaling $1000. The coupon rate of bond A is 0.05 and 0.08 for bond B. A third bond C also exists, with a maturity of two years. Bond C also has a face value of $1000; it has a coupon rate of 11%. The prices of the three bonds are $878.9172, $955.4787, and $1055.419, respectively. Find a portfolio of bonds A, B, and C that would replicate the cash flow structure of bond D, which has a face value of $1000, a maturity of three years, and a coupon rate of 3%.
2. A $1000 face value bond is currently selling at a premium for $1200. The coupon rate of this bond is 12% and it matures in three years. Calculate the following for this bond assuming that its interest payments are made annually:
   a. Annual interest payments
   b. Current yield
   c. Yield to maturity

3. Work through each of the calculations in Problem 2 assuming that interest payments are made semiannually.
4. A life insurance company expects to make payments of $30,000,000 in one year, $15,000,000 in two years, $25,000,000 in three years, and $35,000,000 in four years to satisfy claims of policyholders. These anticipated cash flows are to be matched with a portfolio of the following $1000 face value bonds:

Bond	Current Price	Coupon Rate	Years to Maturity
1	1000	0.10	1
2	980	0.10	2
3	1000	0.11	3
4	1000	0.12	4

   How many of each of the four bonds should the company purchase to exactly match its anticipated payments to policyholders?
5. Find the duration of the following pure discount bonds:
   a. $1000 face value bond maturing in one year currently selling for $900
   b. $1000 face value bond maturing in two years currently selling for $800
   c. $2000 face value bond maturing in three years currently selling for $1400
   d. Portfolio consisting of one of each of the three bonds listed in parts a, b, and c of this problem
6. What is the relationship between the maturity of a pure discount bond and its duration?
7. Find the duration of each of the following $1000 face value coupon bonds assuming coupon payments are made annually:
   a. Three-year 10% bond currently selling for $900
   b. Three-year 12% bond currently selling for $900
   c. Four-year 10% bond currently selling for $900
   d. Three-year 10% bond currently selling for $800
8. Based on duration computations, what would happen to the prices of each of the bonds in Question 7 if market interest rates $(1 + r)$ were to decrease by 10%?
9. What is the duration of a portfolio consisting of one of each of the bonds listed in problem 7?
10. Find durations and convexities for each of the following bonds:
    a. 10% five-year bond selling for $1079.8542 yielding 8%
    b. 12% five-year bond selling for $1000 yielding 12%
11. a. Use the duration (first-order) approximation models to estimate bond value increases induced by changes in interest rates (yields) to 10% for each of the bonds in Problem 10 above.
    b. Use the convexity (second-order) approximation models to estimate bond value increases induced by changes in interest rates (yields) to 10% for each of the bonds in Problem 10 above.

c. Find the present values of each of the bonds in Problem 10 above after yields (discount rates) change to 10%.

12. Consider an example where we can borrow money today for one year at 5%; $y_{0,1} = 0.05$. Suppose that we are able to obtain a commitment to obtain a one-year loan one year from now at an interest rate of 8%. Thus, the one-year forward rate on a loan originated in year equals 8%. According to the pure expectations theory, what is the two-year spot rate of interest $y_{0,2}$?

13. Suppose that the one-year spot rate $y_{0,1}$ of interest is 5%. Investors are expecting that the one-year spot rate one year from now will increase to 6%; thus, the one-year forward rate $y_{1,2}$ on a loan originated in one year is 6%. Furthermore, assume that investors are expecting that the one-year spot rate two years from now will increase to 7%; thus, the one-year forward rate $y_{2,3}$ on a loan originated in two years is 7%. Based on the pure expectations hypothesis, what is the three-year spot rate?

14. Suppose that the one-year spot rate $y_{0,1}$ of interest is 5%. Investors are expecting that the one-year spot rate one year from now will increase to 7%; thus, the one-year forward rate $y_{1,2}$ on a loan originated in one year is 7%. Furthermore, assume that the three-year spot rate equals 7% as well. What is the anticipated one-year forward rate $y_{2,3}$ on a loan originated in two years based on the pure expectations hypothesis?

15. Consider the following four bonds:

F	n	c	P_0
1000	1	0.01	1005
1000	2	0.05	1040
1000	3	0.04	1020
1000	4	0.04	990

Based on the cash flows and prices associated with these bonds, determine the following:

a. Spot rates $y_{0,n}$ for each of four years 1 through 4. These are interest rates on loans originating at time 0.

b. Forward rates $y_{1,t}$ for each of three periods beginning with year 1. These are interest rates on loans originating at time 1.

c. Forward rates $y_{2,t}$ for each of two periods beginning with year 2. These are interest rates on loans originating at time 2.

d. Forward rates $y_{3,n}$ for the period beginning with year 3. This is the interest rates on loans originating at time 3.

16. Bond A, a three-year 7% issue, currently sells for 964.3227. Bond B is a two-year 8% issue currently selling for 1010.031. Bond C is a three-year 6% issue currently selling for 938.4063. Based on this information, answer the following:

a. What are the one-, two-, and three-year spot rates of interest?

b. What are the one- and two-year forward rates on loans originating one year from now?

c. What is the one-year forward rate on a loan originated in two years?

17. Assume that there are two three-year bonds with face values equaling $1000. The coupon rate of Bond A is 0.05 and 0.08 for Bond B. A third Bond C also exists, with a maturity of two years. Bond C also has a face value of $1000; it has a coupon rate of 11%. The prices of the three bonds are $878.9172, $955.4787, and $1055.419, respectively.
- **a.** What are the spot rates implied by these bonds?
- **b.** Find a portfolio of Bonds A, B, and C which would replicate the cash flow structure of Bond D, which has a face value of $1000, a maturity of three years, and a coupon rate of 3%.

18. Suppose we expected that PPP should hold between the United States and Canada. Assume that the exchange rate between U.S. dollars and Canadian dollars is CAD1/USD0.64. That is, one Canadian dollar will purchase USD0.64. If USD2.30 purchases one Big Mac in the United States, how much should a Big Mac cost in Canada?

19. Consider a case involving three currencies concerning cross rates of exchange. Assume that 1.5 Swiss francs are required to purchase 1 U.S. dollar and that 0.64 U.S. dollars are required to purchase 1 Canadian dollar.
- **a.** How many Canadian dollars are required to purchase 10 Swiss francs?
- **b.** Assuming that PPP holds, how many Swiss francs are required to purchase one Big Mac?

20. Assume that $1 will purchase £0.60 and ¥108; that is, 1 U.S. dollar will purchase 0.6 UK pounds and 108 Japanese yen. Assume that goods in the three countries are identically priced after adjusting for currency exchange rates.
- **a.** What is the value of £1 in ¥?
- **b.** What is the value of ¥1 in £?
- **c.** If one ounce of gold costs $300 in the United States, what is its cost in the United Kingdom and in Japan?

21. Consider the following exercise concerned with management of transactions exposure. The Smedley Company has sold products to a Japanese client for ¥15,000,000. Payment is due six months later. Relevant data is as follows:

Spot exchange rate: ¥105/$
Six-month forward exchange rate: ¥104/$
Japanese borrowing interest rate: 9.0%
U.S. borrowing interest rate: 7.0%
Japanese lending interest rate: 7.0%
U.S. lending interest rate: 5.0%
Size of futures contracts: ¥1,000,000
Term to settlement of contracts: six months
Transactions cost on ¥15,000,000 forward contract: $500

Discuss the implications associated with each of the non-options–based methods for managing the transactions exposure risk associated with this extension of credit. Determine which hedging strategy is likely to be optimal and why.

CHAPTER

9

Arbitrage and Hedging With Options

9.1 DERIVATIVE SECURITIES MARKETS AND HEDGING

As discussed earlier, a *derivative security* is simply a financial instrument whose value is derived from that of another security, financial index, or rate. A large number of different types of derivative securities have become increasingly important for management and hedging a variety of different types of risks. There exist a huge variety of derivative securities. Some of the more frequently traded derivatives follow:

Futures contracts, which provide for the transfer of a given asset at an agreed-to price at a future date.
Options, which confer the right but not obligation to buy or sell an asset at a prespecified price on or before a given date. Options are introduced with more detail in the appendix to Chapter 7.
Swaps, which provide for the exchange of one set of cash flows for another set of cash flows.
Hybrids, which combine features of two or more securities (e.g., a convertible bond).

Risk factors frequently hedged with derivatives include, but are not limited to, uncertainties associated with underlying and related asset price movements, interest and exchange rate variability, debtor default, and economy-wide and sector-specific output levels. Markets for explicit insurance policies on such a wide array of risks do not exist largely due to contracting costs. Most insurance policies are fairly standardized (e.g., health, life, and many casualty policies), while customized insurance contracts are expensive and time consuming to write. Businesses must be able to act quickly to manage their risks in this environment of rapid change. Flexibility and liquidity along with low contracting and transactions costs are key to the success of the risk management operations of a firm. An active and efficient market for derivative securities can help firms hedge and manage a variety of different types of risks.

Business firms and individual investors desiring to hedge risks are not the only participants in markets for derivatives. A second type of market participant is the speculator

Financial Trading and Investing
DOI: https://doi.org/10.1016/B978-0-323-90955-6.00009-4

© 2023 Elsevier Inc. All rights reserved.

who takes a position in a security based on his expectation regarding future price movement. Although the speculator is essentially concerned with his own trading profits, he plays an important role in maintaining liquidity in derivative markets, affording business and individual investors the opportunity to hedge risks quickly and efficiently. The speculator is often the counterparty to a hedger's trade, selling or purchasing derivatives as required by hedgers.

The arbitrageur, who exploits situations where derivatives are mispriced relative to one another, not only provides additional liquidity to derivative markets, but plays an important role in their pricing. By constantly seeking price misalignments for a variety of types of securities, and by understanding the payoffs of securities relative to one another, arbitragers help ensure that derivative securities are fairly priced. This activity reduces price volatility and uncertainty faced by hedgers. In addition, the presence of arbitrageurs provides liquidity to other market participants.

As discussed in Chapter 2, derivative securities are traded in the United States either on exchanges or over the counter (OTC) markets. Substantial market interest is required for exchange listing, whereas securities with smaller followings or even customized contracts can be traded OTC, including trading between banks and other major participants. The role of the derivatives dealer is essentially the same as that for other security dealers: to facilitate transactions for clients at competitive prices. Derivative dealers match counterparties for derivative contracts, act as counterparty for many of their own custom contracts, and provide an array of support services including expert advice and carefully engineered customized risk management products. It is necessary that the dealer providing full support services have a proper understanding of his client and the client's business and needs. Since some clients do not have an understanding of the technical terms used in the industry, the dealer must be an effective communicator. It is equally important for the dealer to understand the nature of the securities with which he deals and how serving as a market maker for derivatives affects the risk structure of his employer. This understanding usually requires strong analytical skills.

Many stock options in the United States and Europe are traded on exchanges. The largest U.S. stock options exchanges are the Chicago Board Options Exchange and the NASDAQ OMX Group. The NYSE operates the options markets of the former American Exchange and the International Stock Exchange maintains an options market. Options are also traded on numerous commodities, futures contracts, currencies, and other financial instruments such as Treasury instruments and index contracts.

Options can be classified into either the European variety or the American variety. European options may be exercised only at the time of their expiration; American options may be exercised any time before and including the date of expiration. Most option contracts traded in the United States (and Europe as well) are of the American variety. American options can never be worth less than their otherwise identical European counterparts. In fact, because most call options have time value in addition to their intrinsic or exercise value (calls on stocks that go ex-dividend before the call expires can be an important exception to this), we usually do not exercise American calls before exercise. This means that we can often value American call options (on non-dividend−paying stock) as though they are European calls.

9.2 PUT–CALL PARITY

First, since the call (put) owner has the right to buy (sell) the underlying stock at a price of X, the terminal payoff functions for calls and puts are written as follows:

$$c_T = \text{MAX}[0,\ S_T - X] \tag{9.1}$$

$$p_T = \text{MAX}[0,\ X - S_T] \tag{9.2}$$

If we subtract Equation (9.2) from Equation (9.1) we obtain the terminal value *put–call relation*:

$$c_T - p_T = \text{MAX}[0,\ S_T - X] - \text{MAX}[0,\ X - S_T] = S_T - X \tag{9.3}$$

A slight rewrite of this terminal put–call relation allows us to write the terminal or exercise value a put given the terminal value of a call with identical exercise terms:

$$p_T = c_T + X - S_T \tag{9.4}$$

Since the terminal value of a put is always given by Equation (9.4), the time-zero value of a put must be given by Equation (9.5):

$$p_0 = c_0 + Xe^{-r_f T} - S_0 \text{ in continuous time};\ p_0 = c_0 + X/(1+r_f)^T - S_0 \text{ in discrete time} \tag{9.5}$$

This arbitrage-free relationship allows us to value a put based on the price knowledge of a call with exactly the same exercise terms. The put–call parity function holds regardless of the stochastic process generating stock prices, but assumes that the underlying stock pays no dividends during the lives of the options.

Illustration: Hedging With a Call, a Put and a Collar

Suppose that an investor has 1000 shares of stock that are currently worth \$50 per share each. The investor wishes to lock in his gains on the stock, but does not want to sell at this time for tax reasons. The current two-year riskless return rate is 5%. The investor could write two-year calls on these shares, with an exercise price of \$50 per share, receiving \$8 for each share that he owns. That is, he could receive \$8 per share by selling covered calls. He could also purchase two-year puts on each share which would enable him to eliminate downside risk. What are the potential hedging strategies available to the investor? What are the costs and benefits of each?

First, the investor could simply write covered calls. This strategy offers no downside risk protection, but it does provide the investor with \$8 (or \$8000 total) in immediate cash flows. The strategy also limits the investor's upside potential on the shares, locking out any gains should the share price rise above \$50. This is because the investor would be obliged to sell the stock to the owner of the call should its price rise above \$50 per share. The owner of the call would not exercise this option if the share price remains below \$50. While the covered call strategy does reduce portfolio volatility, it is more significant to the investor as a means of generating short-run income at the expense of potential stock profits.

A second potential strategy for the investor is to purchase puts. We can use put–call parity to value these puts. Since the riskless rate is $r_f = 0.05$, the exercise price is $X = \$50$, the current share price is $S_0 = \$50$, and the $T =$ two-year call price is $c_0 = \$8$, we can use Equation (9.5) to value the put:

$$p_0 = 8 + 50e^{-0.05 \times 2} - 50 = 3.24$$

Each put will protect the investor from downside stock price movements below \$50 per share, and, based on the data given, a fair price for each put is \$3.24. Thus, for \$3.24, the investor can lock in a minimum selling price of \$50 for each share. The puts will not require the investor to give up any gains above \$50.

What if the investor objects to paying \$3.24 to eliminate downside price risk? Another possibility is to simultaneously purchase a put and write a call, in effect financing the cost of the put with proceeds from selling the call. One such strategy is to create a *collar*, which is a package consisting of a long position in a put and a short position in a call. Here, if the investor purchases a put for \$3.24 and sells a call for \$8, he nets \$4.76 per share. If the share price drops below \$50 in two years, he puts his shares to the put writer for \$50. If the share price rises above \$50 in two years, his shares are called away from him at \$50 per share. Thus, for a net cash flow today of \$4.76, the investor gives up all potential profits and losses on his shares, locking or "collaring" in a price of \$50 today.

9.3 OPTIONS AND HEDGING IN A BINOMIAL ENVIRONMENT

The *binomial option pricing model* is based on the assumption that the underlying stock follows a binomial return generating process. This means that for any period during the life of the option, the stock's value will be only one of two potential constant values. For example, the stock's value will be either u (multiplicative upward movement) times its current value or d (multiplicative downward movement) times its current value.

Consider a stock currently selling for 100 and assume for this stock that u equals 1.2 and d equals 0.8. The stock's value in the forthcoming period will be either 120 (if outcome u is realized) or 80 (if outcome d is realized). Suppose that there exists a European call trading on this particular stock during this one-time period model with an exercise price of 90. The call expires at the end of this period when the stock value is either 120 or 80 as in Figure 9.1. Thus, if the stock were to increase to 120, the call would be worth 30 ($c_u = 30$), since one could exercise the call by paying 90 for a stock which is worth 120. If the stock value were to decrease to 80, the value of the call would be zero ($c_d = 0$) since no one would wish to exercise by paying 90 for shares that are worth only 80. Furthermore, suppose that the current riskless return rate is 0.10. Based on this information, we should be able to determine the value of the call.

Note that we have not specified probabilities of a stock price increase or decrease during the period prior to option expiration. Nor have we specified a discount rate for the option or made inferences regarding investor risk preferences. We will value this call based on the fact that during this single time period, we can construct a riskless hedge

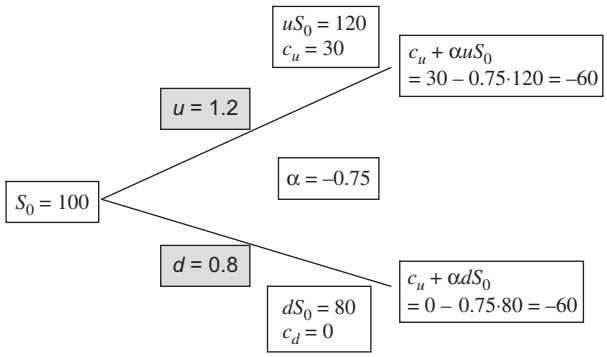


FIGURE 9.1 One-time-period binomial model.

portfolio consisting of a position in a single call and offsetting positions in α shares of underlying stock. This means that by purchasing a single call and by selling α shares of stock, we can create a portfolio whose value is the same regardless of whether the underlying stock price increases or decreases. Let us first define the following terms:

S_0 = Initial stock value
u = Multiplicative upward stock price movement
d = Multiplicative downward stock price movement
c_u = MAX[0,$uS_0 - X$]; value of call if stock price increases
c_d = MAX[0,$dS_0 - X$]; value of call if stock price decreases
α = Hedge ratio
r_f = Riskless return rate
X = Exercise price of the call

In our numerical example offered above, we use the following to determine the value of the call in the binomial framework:

$S_0 = 100$
$u = 1.2$
$d = 0.8$
$c_u = 30$
$c_d = 0$
$r_f = 0.10$
$X = 90$

Our first step in determining the value of the call is to determine α, the *hedge ratio*. The absolute value of the hedge ratio defines the number of shares of stock that must be sold (or short sold) for each long call position in order to maintain a riskless portfolio. This riskless portfolio will be comprised of one call option along with a short position in $-\alpha$ shares of stock, actually, the absolute value of α shares of stock, since alpha is negative. In this case, the riskless hedge portfolio made up of one call option and α shares of stock will have the same value whether the stock price increases to uS_0 or decreases to dS_0 (see Figure 9.1).

If we were to purchase one call and sell $-\alpha$ shares of stock, our riskless hedge condition is given as follows:

$$c_u + \alpha u S_0 = c_d + \alpha d S_0 \tag{9.6}$$
$$30 + \alpha \cdot 120 = 0 + \alpha \cdot 80$$

We defined the absolute value of α to be the number of shares to sell for every call purchased. This value is known as a hedge ratio. By maintaining this hedge ratio, we maintain our hedged portfolio. Solve for the hedge ratio α as follows:

$$\alpha = \frac{c_u - c_d}{S_0(d-u)}$$
$$\alpha = \frac{30 - 0}{100(0.8 - 1.2)} = -0.75 \tag{9.7}$$

In our example, $\alpha = -0.75$, meaning that the investor should sell 0.75 shares of underlying stock for each call option that he purchases to maintain a riskless portfolio. If the investor shorts 0.75 shares for each call purchased, the riskless portfolio will be worth -60 regardless of whether the stock rises to 120 or drops to 80 (see Figure 9.1). Similarly, if the investor purchased 0.75 shares for each call sold or written, the riskless portfolio will be worth 60 regardless of whether the stock rises to 120 or drops to 80. Thus, as long as the investor can confirm that there are only two potential outcomes for the underlying stock price, and determine what these prices are, he can create a hedge or riskless portfolio comprised of the option and α shares of the underlying stock. Since this hedge portfolio is riskless, it must earn the riskless rate of return for arbitrage opportunities to be avoided:

$$c_u + \alpha u S_0 = c_d + \alpha d S_0 = (c_0 + \alpha S_0)(1 + r_f) \tag{9.8}$$

Thus, our problem now is to value the call, c_0. From here, we can work equally well either with outcome u or outcome d; since it makes no difference we can choose outcome d. Note that the time-zero option value c_0 can be solved for by rearranging Equation (9.8). If Equation (9.8) does not hold, or if the current price of the option is inconsistent with Equation (9.8), a riskless arbitrage opportunity will exist. Thus, we will rewrite Equation (9.8) to solve for the zero NPV condition that eliminates positive profit arbitrage opportunities:

$$c_0 + \alpha S_0 = \frac{c_d + \alpha d S_0}{(1 + r_f)} \tag{9.9}$$

It is now quite simple to solve for the call value c_0 by rewriting Equation (9.9):

$$c_0 = \frac{-(1 + r_f)\alpha S_0 + c_d + \alpha d S_0}{(1 + r_f)}$$
$$c_0 = \frac{(1 + 0.10) \cdot 0.75 \cdot 100 + 0 - 0.75 \cdot 0.8 \cdot 100}{(1 + 0.10)} = 20.4545 \tag{9.10}$$

Equation (9.10) is quite appropriate for evaluating a European call in a one-time period binomial framework. That is, in the model presented thus far, share prices can either increase or decrease once by a prespecified amount or percentage. Thus, there are only two potential prices that the stock can assume at the expiration of the stock. This binomial option pricing model can easily be extended to cover as many potential outcomes and time periods as necessary for a particular scenario.

Arbitrage in the One-Time Period Case

We found the arbitrage-free value of the call in the above example to be 20.4545. What if its price in the marketplace were instead 19? In this scenario, since the call is undervalued, we will purchase it, along with taking a short position in 0.75 shares of the underlying stock at \$100. This initial investment will be $-\$56$, which we will loan at the riskless rate of 10%. Our initial and time-one cash flows are computed as follows:

$$\text{Time-zero cash flow: } 1\times -19.00 - 0.75\times -100 - 56 = 0$$
$$\text{Time-one cash flow: } 1\times 30.00 - 0.75\times 120 + 56(1.1) = 1\times 0 - 0.75\times 80 + 56(1.1) = 1.6$$

Thus, the time-zero cash flow nets to zero. If the stock price rises to 120 in time one, the call pays \$30, the short position requires \$90 to settle, and the loan repays \$61.60. If the stock price drops to 80 in time one, the call pays 0, the short position requires \$60 to settle and the loan repays \$61.60. Regardless, the time-one net cash flow is \$1.60, representing a time-one arbitrage profit equal to 1.6, resulting from a time zero investment of zero.

Extending the Binomial Model to Two Periods

As we stated above, Equation (9.10) is appropriate for evaluating a European call in a one-time-period framework. That is, in the model presented thus far, share prices can either increase or decrease once by a prespecified percentage. Thus, there are only two potential prices that the stock can assume at the expiration of the stock. Our next step in the development of a more realistic model is to extend the framework to two time periods. One complication is that the hedge ratio only holds for the beginning of the first time period. After this period, the hedge ratio must be updated to reflect price changes and movement through time. Thus, our first step in extending the model to two time periods is to substitute for the hedge ratio based on Equation (9.10):

$$c_0 = \frac{(1+r_f)\dfrac{(c_u - c_d)}{S_0(u-d)}S_0 + c_d - \dfrac{(c_u - c_d)}{S_0(u-d)}dS_0}{(1+r_f)} \tag{9.11}$$

The next two steps of our development are to simplify Equation (9.11):

$$c_0 = \frac{\dfrac{(1+r_f)(c_u - c_d) + c_d(u-d) - d(c_u - c_d)}{(u-d)}}{(1+r_f)} \tag{9.12}$$

$$c_0 = \frac{c_u \dfrac{(1+r_f)-d}{(u-d)} + c_d \dfrac{u-(1+r_f)}{(u-d)}}{(1+r_f)} \tag{9.13}$$

Assume for the moment that r_f can be regarded as a riskless discount rate and that investors will discount cash flows derived from the call based on this riskless rate; that is, for the moment, assume that investors are risk neutral. This assumption is reasonable if investors investing in options behave as though they are risk neutral; in fact they will evaluate options as though they are risk neutral because they can eliminate risk by setting appropriate hedge ratios. Their extent of risk aversion will already be reflected in the current value that they associate with underlying stock given its potential future values. Note that Equation (9.13) defines the current call value in terms of the two potential call prices cu and cd and this riskless return rate. Since it is reasonable to assume that investors behave towards options as though they are risk neutral, the numerator of Equation (9.13) might be regarded as an expected cash flow. Hence, the terms which c_u and c_d are multiplied by might be regarded as probabilities. Define p to be the probability that the risk neutral investor associates with the stock price changing to uS_0 and $(1 - p)$ as the probability p that the stock price changes to dS_0:

$$p = \frac{(1+r_f)-d}{(u-d)} \qquad (1-p) = \frac{u-(1+r_f)}{(u-d)} \tag{9.14}$$

We simplify Equation (9.13) by substituting in Equations (9.14):

$$c_0 = \frac{c_u p + c_d(1-p)}{(1+r_f)} \tag{9.15}$$

Equation (9.15) represents a slightly simplified version of the one-time period option pricing model. However, it is easily extended to a two-period model if we assume that u and d are constant over the periods. Let c_{u^2} be the call's expiration value after two time periods assuming that the stock's price has risen twice; c_{d^2} is the value of the call assuming the stock price declined twice during the two periods and c_{ud} is the value of the call assuming the stock price increased once and decreased once during these two periods. Thus, our two-time-period model becomes:

$$c_0 = \frac{c_{u^2}p^2 + 2p(1-p)c_{ud} + (1-p)^2 c_{d^2}}{(1+r_f)^2} \tag{9.16}$$

Illustration: Two-Time-Period Hedges

Now, we will extend our illustration above from a single period to two periods, each with a riskless return rate equal to 0.10. As before, the stock currently sells for 100 and will change to either 120 or 80 in one time period ($u = 1.2$, $d = 0.8$). However, in the second period, the stock will change a second time by a factor of either 1.2 or 0.8, leading to

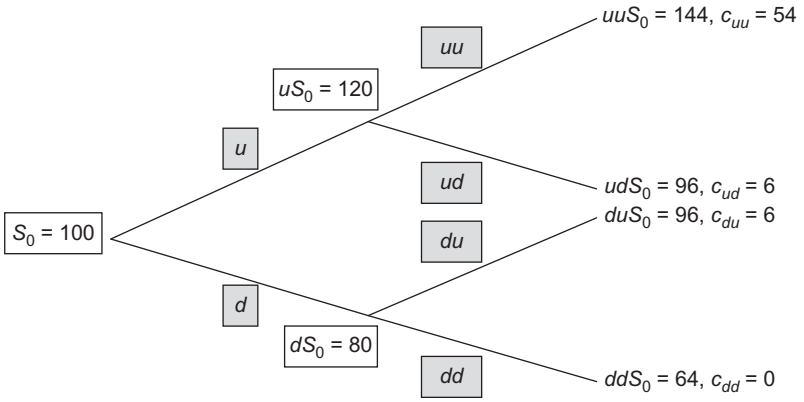


FIGURE 9.2 Two-time-period binomial model.

potential values of either 144 (up then up again), 96 (up once and down once), or 64 (down twice). The lattice (tree) associated with this stock price process is depicted in Figure 9.2. Call option prices are also given for the second period. The probability of upward and downward movements are computed with the Equation (9.14) set as follows:

$$p = \frac{((1+0.1)-0.8)}{(1.2-0.8)} = 0.75 \qquad (1-p) = \frac{(1.2-(1+0.05))}{(1.2-0.8)} = 0.25$$

The time-zero call value in this two-period binomial framework is computed with Equation (9.16) as follows:

$$c_0 = \frac{(144-90)\times 0.75^2 + 2\times 0.75(1-0.75))\times(96-90) + (1-0.75)^2\times 0}{(1+0.10)^2} = 26.96281$$

Because the call is priced relative to the stock and riskless asset, any deviation of the market price of the call from this theoretical binomial price creates an arbitrage opportunity for investors, just as in the one-time period illustration above.

We calculate the hedge ratio for time zero to be -0.75 from Equation (9.7) and the hedge ratio in time one to be either -0.875 or -1, depending on whether the share price increases or decreases in the first period:

$$\alpha_0 = \frac{30-0}{100(0.8-1.2)} = -0.75$$

$$\alpha_u = \frac{54-6}{120(0.8-1.2)} = -1 \quad \alpha_d = \frac{6-0}{80(0.8-1.2)} = -0.1875$$

Thus, the investor will hedge by shorting 0.75 shares for each long call in the first period, and will update his hedge ratio to either 1 or 0.1875 short shares in the second period. Also, note the change in time perspectives for u and d in the second period after one period has elapsed.

Extending the Binomial Model to *n* Time Periods

The binomial model is easily extended to n time periods (evenly split within the interval T) by implementing the binomial theorem to obtain the following[1]:

$$c_0 = \frac{\sum_{j=0}^{n} \frac{n!}{j!(n-j)!} p^j (1-p)^{n-j} \text{MAX}\left[0, u^j d^{n-j} S_0 - X\right]}{(1+r_f)^n} \tag{9.17}$$

Computing Equation (9.17) might be simplified by using Equation (9.18), where a is determined by Equation (9.19)[2]:

$$c_0 = \frac{\sum_{j=a}^{n} \frac{n!}{j!(n-j)!} p^j (1-p)^{n-j} \left[u^j d^{n-j} S_o - X\right]}{(1+r_f)^n} \tag{9.18}$$

$$a = \text{MAX}\left[INT\left(\frac{\ln\left(\frac{X}{S_0 d^n}\right)}{\ln\left(\frac{u}{d}\right)} + 1\right), 0\right] \tag{9.19}$$

The binomial option pricing model prices calls and puts relative to the pricing of underlying securities and the riskless return. That is, the binomial model prices an option against a portfolio comprising the underlying security and the riskless asset. Any deviation in the market prices from their binomial prices suggests one or both of the following:

1. The binomial option pricing model does not apply to this market.
2. There is an arbitrage opportunity in this market.

Illustration: Three Time Periods

Now, we will extend our illustration above from two periods to three, each with a riskless return rate equal to 0.10. As we see in Figure 9.3, by the third period, potential stock values are $1.2^3 \times 100 = 172.8$, $1.2^2 \times 0.8 \times 100 = 115.2$, $0.8^2 \times 1.2 \times 100 = 76.8$, or

[1]See Appendix 9.A for details on the application of this model and its relationship to the Black-Scholes model. See also Table 9.2.

[2]See Appendix 9.A for a derivation of Equation (9.19). a is determined by Equation (9.19) and is the first positive integer where $u^j d^{n-j} S_0 > X$. That is, the minimum number of up-jumps required for exercise of the call option equals a. Any smaller number of stock up-jumps produces a terminal call value equal to zero, and need not be further considered.

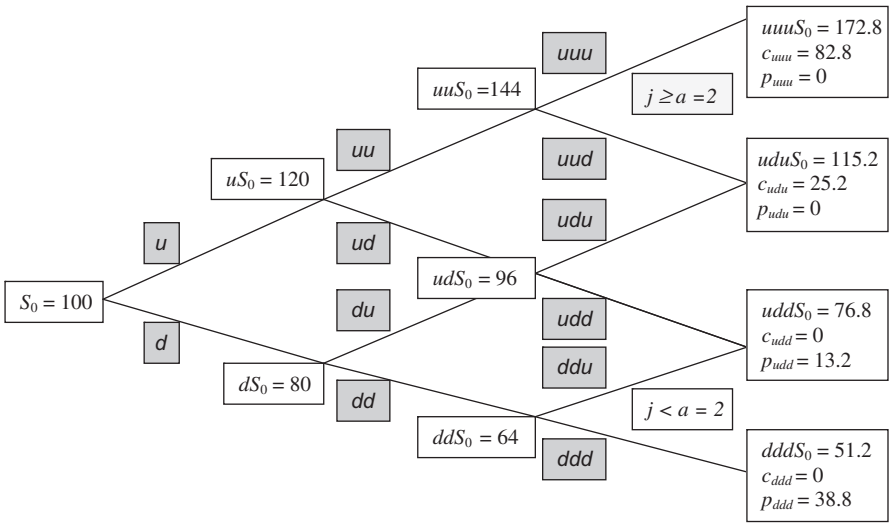


FIGURE 9.3 Three-time-period binomial model.

$0.8^3 \times 100 = 51.2$. The time-zero call value in this three-period binomial framework is computed with Equations (9.19) and (9.18) as follows:

$$a = \text{MAX}\left[INT\left(\frac{\ln\left(\frac{90}{100 \times 0.8^3}\right)}{\ln\left(\frac{1.2}{0.8}\right)} + 1 \right),\ 0 \right] = 2$$

$$c_0 = \frac{(3 \times 0.75^2 \times 0.25) \times (1.2^2 \times 0.8 \times 100 - 90) + (1 \times 0.75^3) \times (1.2^3 \times 100 - 90)}{(1+0.10)^3}$$

$$= \frac{(0.421875 \times 25.2) + (0.421875 \times 82.8)}{1.331} = \frac{10.63125 + 34.93125}{1.331} = 34.23178$$

Put–call parity still applies[3]:

$$p_0 = 34.23178 + 90/(1.1)^3 - 100 = 1.85$$

Obtaining Multiplicative Upward and Downward Movement Values

One apparent difficulty in applying the binomial model as it is presented above is in obtaining estimates for u and d that are required for p; all other inputs are normally quite easily obtained. There are several methods that are used to obtain parameters for the

[3]We will discount the exercise money with a discrete discount function since the binomial model is a discrete time model.

binomial method from the actual security returns generating process. For example, following Cox, Ross, and Rubinstein (1979), derive the following to estimate probabilities of an uptick p and downtick $(1 - p)$:[4]

$$p = \frac{e^{r_f} - d}{u - d} \quad (1 - p) = \frac{u - e^{r_f}}{u - d} \tag{9.20}$$

Cox et al. also propose the following to estimate u and d in the binomial approximation to the Wiener process, where σ is the standard deviation of stock returns:

$$u = e^{\sigma} \quad d = \frac{1}{u} \tag{9.21}$$

or, if n and T differ from 1:

$$u = e^{\left(\sigma \times \sqrt{\frac{T}{n}}\right)} \quad d = \frac{1}{u} \tag{9.22}$$

Suppose, for example, that for a particular Wiener process, $\sigma = 0.30$ and $r_f = 0.05$. Using Equations (9.20) and (9.22), we estimate p, u, and d for a single-time-period binomial process as follows:

$$u = e^{0.3} = 1.3498588$$

$$d = \frac{1}{u} = 0.7408182$$

Binomial Model: An Illustration

Suppose that we wished to evaluate a call and a put with an exercise price equal to \$110 on a share of stock currently selling for \$100. The underlying stock-return standard deviation equals 0.30 and the current riskless return rate equals 0.05. If both options are of the European variety and expire in six months, what are their values?

First, we will compute the call's value using the binomial model. We will vary the number of jumps in the model as the example progresses. First, let $n = 1$ and use Equations (9.20) and (9.22) to compute p, u, and d:

$$u = e^{0.3 \times \sqrt{0.5}} = 1.2363111$$

$$d = \frac{1}{1.2363111} = 0.8088578$$

[4]See Appendix 9.A for derivations of p, u, and d. Regardless, notice the somewhat minor deviation from the probability estimates given by the Equation (9.14) set. The difference between the Equations (9.14) and (9.20) sets is that Equation (9.20) allows interest (r_f) to be continuously compounded, whereas Equation (9.14) is based on discrete compounding. This distinction is not important for our purposes here.

$$p = \frac{e^{r_f T} - d}{u - d} = 0.5063881$$

Thus, in a risk-neutral environment, there is a 0.5064 probability that the stock price will increase to 123.63 and a 0.4936 probability that the stock price will be 80.88678. Similarly, in a risk-neutral environment, there is a 0.5064 probability that the call will be worth 13.63; therefore, its current value is $6.73 = 0.5064 \times 13.63 \times e^{-0.05 \times 0.5}$. The call value is determined by the binomial model as follows where $n = a = 1$:

$$c_0 = \frac{\sum_{j=a}^{n} \frac{n!}{j!(n-j)!} p^j (1-p)^{n-j} \left[u^j \times d^{n-j} S_o - X\right]}{(1+r_f)^T}$$

$$c_0 = \frac{0.506388^1 \times 0.4936119^{1-1} \times [1.236311^1 \times 0.8088578^{1-1} \times 100 - 110]}{(1+0.05)^5} = 6.73$$

We can also use the binomial model to value the put with identical exercise terms on the underlying stock[5]:

$$p_0 = \frac{\sum_{j=0}^{a-1} \frac{n!}{j!(n-j)!} p^j (1-p)^{n-j} \left[X - u^j d^{n-j} S_o\right]}{(1+r_f)^T}$$

$$p_0 = \frac{0.506388^0 \times 0.4936119^1 \times [110 - 1.23631^0 \times 0.8088578^1 \times 100]}{(1+0.05)^5} = 14.08$$

Now, revise the illustration by dividing the single six-month interval into two three-month intervals; that is, $n = 2$. We will now use a two-period binomial model to evaluate calls and puts on this stock. First, we use Equations (9.20) and (9.22) to compute p, u, and d[6]:

$$u = e^{\left(0.3 \times \sqrt{\frac{0.5}{2}}\right)} = 1.1618342$$

$$d = \frac{1}{1.1618342} = 0.8607079$$

$$p = \frac{e^{r_f T/n} - d}{u - d} = 0.5043415$$

[5]Recall that put–call parity, demonstrated in the Options and Hedging in a Binomial Environment section, can be used to value puts. We verify this for this example by using Equation (9.3) as follows: $14.08 = 6.73 + 110e^{-0.05 \times 0.5} - 100$.

[6]When there are multiple jumps per period ($n > 1$), and/or when T does not equal 1, $p = \frac{e^{r_f(T/n)} - d}{u - d}$ and $(1 - p) = \frac{u - e^{r_f(T/n)}}{u - d}$.

Thus, there is a 0.5043^2 probability that the stock price will increase to 134.98, a 0.5 probability that the stock price will remain unchanged at 100, and a 0.4957^2 probability that the stock price will decline to 74.08. Thus, there is a 0.2543 probability that the call will be exercised, in which case, it will be worth 24.98. Therefore, the call's current value is $6.20 = 0.2543 \times 24.98 \times e^{-0.05 \times 0.5}$. Call and put values are determined by the binomial model as follows where $n = a = 2$:

$$c_0 = \frac{\sum_{j=a}^{n} \frac{n!}{j!(n-j)!} p^j (1-p)^{n-j} \left[u^j d^{n-j} S_o - X\right]}{(1+r_f)^T}$$

$$c_0 = \frac{0.5043^2 \times 0.4957^{2-2} \times [1.16185^2 \times 0.8607^{2-2} \times 100 - 110]}{(1+0.05)^{0.5}} = 6.20$$

$$p_0 = \frac{\sum_{j=0}^{a} \frac{n!}{j!(n-j)!} p^j (1-p)^{n-j} \left[X - u^j d^{n-j} S_o\right]}{(1+r_f)^T}$$

$$p_0 = \frac{0.5043^{2-2} \times 0.4957^2 \times [110 - 1.1618^{2-2} \times 0.8607^2 \times 100] + 2 \times 0.5043 \times 0.4957 \times [110-100]}{(1+0.05)^{0.5}}$$

$$= 6.20 + 110e^{-0.05 \times 5} - 100 = 13.48$$

As the six-month period is divided into more and finer subintervals, the values of the call and put will approach their Black-Scholes values. Table 9.1 extends this example to more than two subintervals.

9.4 THE GREEKS AND HEDGING IN A BLACK-SCHOLES ENVIRONMENT

As the number of trials n in a binomial distribution approaches infinity, the binomial distribution approaches the normal distribution. Black and Scholes (see Appendix 9.B) provide a derivation for an option pricing model based on the assumption that the natural logs of stock price relatives will be normally distributed.[7] We introduced the Black-Scholes model in the appendix to Chapter 7. The assumptions on which the *Black-Scholes options pricing model* and its derivation are based are as follows:

1. There are no restrictions on short sales of stock or writing of call options.
2. There are no taxes or transactions costs.
3. There exists continuous trading of stocks and options.
4. There exists a single constant riskless interest rate that applies for both borrowing and lending.

[7]The stock price relative for a given period t is defined as $S_t \div S_{t-1}$. Thus, the log of the stock price relative is defined as $\ln[S_t \div S_{t-1}]$.

TABLE 9.1 Convergence of Binomial Model to Black-Scholes Model

n	c_0	p_0_
1	6.73	14.02
2	6.20	13.48
3	5.47	12.72
4	6.04	13.30
5	5.18	12.44
6	5.91	13.17
7	5.43	12.68
8	5.81	13.06
9	5.57	12.82
10	5.73	12.98
50	5.63	12.89
100	5.59	12.85
∞	5.59	12.85
Volatility	$\sigma = 0.30$	
Riskless rate	$r_f = 0.05$	
Exercise price	$X = 110$	
Initial stock price	$S_0 = 100$	
Term to expiration	$t = 0.5$	

5. The range of potential stock prices is continuous. This means that a stock's price can take on any real value.
6. The underlying stock will pay no dividends during the life of the option.
7. The option can be exercised only on its expiration date; that is, it is a European option.
8. Shares of stock and option contracts are infinitely divisible.
9. Stock prices follow an Îto process; that is, they follow a continuous time-random walk in two-dimensional continuous space. This simply means that stock prices are randomly distributed and can take on any positive value at any time.

From an applications perspective, one of the more useful aspects of the Black-Scholes model is that it only requires five inputs. All of these inputs, with the exception of the variance of underlying stock returns, is normally quite easily obtained[8]:

1. The current stock price (S_0): Use the most recent quote.
2. The variance of returns on the underlying stock (σ^2): Several methods for estimating stock volatilities were discussed in Chapter 7.

[8]These five inputs are the only ones necessary if the assumptions underlying the model hold.

3. The exercise price of the option (X) is given by the contract.
4. The time to maturity of the option (T) is given by the contract.
5. For the risk-free return rate (r_f), use a Treasury instrument rate with an appropriate term to maturity.

It is important to note that the following less easily obtained factors are not required as model inputs:

1. The expected or required return on the stock or option.
2. Investor attitudes toward risk.

If the assumptions given above hold, the Black-Scholes model specifies that the value of a call option is as follows:

$$c_0 = S_0 N(d_1) - \frac{X}{e^{r_f T}} N(d_2) \tag{9.23}$$

$$d_1 = \frac{\ln(S_0/X) + (r_f + 0.5\sigma^2)T}{\sigma\sqrt{T}} \tag{9.24}$$

$$d_2 = d_1 - \sigma\sqrt{T} \tag{9.25}$$

where $N(d^*)$ is the cumulative normal distribution function for (d^*).[9]

Black-Scholes Model Sensitivities

Options can be very volatile instruments, and hedging is crucial for traders' management of options portfolios. Sources of sensitivity, or the "Greeks," are key to this management process. The Black-Scholes model can provide much useful information concerning options and underlying shares. First, assuming that investors behave as though they are risk neutral and that other Black-Scholes assumptions hold, $N(d_2)$ might be interpreted as the probability that the stock price will be sufficiently high at the expiration date of the option to warrant its exercise. Thus, $N(d_2)$ might be regarded as the probability that investors assign to S_T exceeding X; that is, the probability at exercise at time T implied by the call at its current market price c_0 is $N(d_2)$. A second particularly useful point is that $N(d_1)$ might be interpreted as a hedge ratio, although it must be updated at every period. This means that $N(d_1)$ can be interpreted as the number of shares to short for every purchased call, but, as the option's time to maturity diminishes, and as share prices change, this hedge ratio will change. In addition, $N(d_1)$ is the sensitivity of the option value to changes in the stock price. This sensitivity means that the option's value will change by $N(d_1)$ for each unit of change in the underlying stock price. In the finance industry, this value is often referred to as the call *delta*.

Option traders find it very useful to know how the values of their option positions will change as the various factors used in the pricing model change. Knowledge of these sensitivities (sometimes called *Greeks*) is particularly useful to investors holding portfolios of options and can often be extended to weighted averages for the portfolios.

[9]See the appendices to Chapter 7 for an introduction to the Black-Scholes model.

For example, we mentioned above that the sensitivity of the call's value to the stock's price is Delta:[10]

$$\frac{\partial c}{\partial S} = N(d_1) > 0 \qquad Delta \quad \Delta \tag{9.26}$$

Thus, a small increase in the value of the underlying stock would lead to approximately $N(d_1)$ times that increase in the price of the call option. A call investor can hedge his portfolio risk associated with infinitesimal share price changes by shorting $N(d_1)$ shares of underlying stock for each purchased call option. Thus, the hedged portfolio would consist of an offsetting position of $N(d_1)$ underlying shares for each call option in the portfolio.

In addition, the trader can offset this delta risk with multiple positions in options and the underlying stock. For example, if the trader was long a single call option with a delta of 0.5, this risk could be neutralized with offsetting positions in a portfolio of stock and one or more other option contracts on that stock. Thus, in our example, the stock delta could be offset with a portfolio whose weighted average delta was -0.5, comprised of a single short position in a call option with a delta equal to 0.25 and a short position in 0.25 shares of the underlying stock. Any portfolio whose weighted average delta is 0.5 could be comprised to offset the original call option delta risk.

There is a major caveat to the application of this hedge ratio. In addition to the Black Scholes assumptions, this sensitivity or delta measure worsens as the change in the stock price increases. Furthermore, a large change in the stock price will lead to a change in the value of the *delta neutral portfolio* because delta is based on a derivative or "point slope," and applies only to infinitesimal changes in the stock price. Thus, because this delta is based on a partial derivative with respect to the share price, it holds exactly only for an infinitesimal change in the share price; it holds only approximately for finite changes in the share price. This delta only approximates the change in the call value resulting from a change in the share price because any change in the price of the underlying shares would lead to a change in the delta itself:

$$\frac{\partial^2 c}{\partial S^2} = \frac{\partial \Delta}{\partial S} = \frac{\partial N(d_1)}{\partial S} = \frac{e^{\left(\frac{-(d_1^2)}{2}\right)}}{S\sigma\sqrt{2\pi T}} > 0 \qquad Gamma \quad \Gamma \tag{9.27}$$

This change in delta resulting from a change in the share price is known as *gamma*. Since gamma is positive, an increase in the share price will lead to an increase in delta. However, again, this change in delta resulting from a finite share price change is only approximate. Each time the share price changes, the investor must update his portfolio to reflect the changing delta.

Gamma is very useful to traders that want to hold delta neutral portfolios. With a fairly large number of puts and calls on the same underlying stock to choose from, traders will be able to discern an infinity of portfolios (in theory with an infinity of potential stock price changes) that maintain delta neutrality. That is, with a single option and shares of its underlying stock, there will normally be a single hedge ratio that provides for portfolio

[10]Applying the Chain Rule when finding this derivative yields two additional terms that cancel each other.

delta neutrality. This delta neutrality means that fairly small changes in the underlying stock's price will have a negligible impact on the value of the portfolio of shares and options. With two different options contracts, for example, contracts with different exercise prices, there may be an infinity of portfolios that maintain delta neutrality. However, there will probably be only one of these portfolios that simultaneously maintains delta and gamma neutrality (select portfolio weights appropriately because all gammas are positive). This delta-gamma neutral portfolio will normally minimize adjustments necessary to accommodate near-term future stock price changes in the hedged portfolio. At least two different options contracts along with underlying shares will be needed to maintain both delta and gamma neutrality. With two different options contracts and underlying shares, there will normally be a single combination of portfolio weights that simultaneously maintains delta and gamma neutrality; more than two different options contracts and underlying shares will allow for an infinite number of delta/gamma neutral portfolios.

Since each call and put option has a date of expiration, calls and puts are said to amortize over time. As the date of expiration draws nearer (T gets smaller), the value of the European call (and put) option might be expected to decline as indicated by a positive *theta*:

$$\frac{\partial c}{\partial T} = r_f X e^{-r_f T} N(d_2) + S\frac{\sigma}{\sqrt{T}}\frac{e^{\left(\frac{-(d_1^2)}{2}\right)}}{2\sqrt{2\pi}} > 0 \qquad Theta \qquad \theta \tag{9.28}$$

Many traders seek to maintain portfolios that are simultaneously neutral with respect to delta, gamma, and theta. Be certain that the units of time for theta are consistent with those used in the calculation of the option value. Some option traders calculate option prices based on numbers of years (fractions of years) prior to option expiry, but prefer to calculate theta in days. This is acceptable, but divide the annual theta by 365 (or the number of trading days in the year) to obtain the daily theta.

Vega, which actually is not a Greek letter, measures the sensitivity of the option price to the underlying stock's standard deviation of returns. One might expect the call option price to be directly related to the underlying stock's standard deviation:

$$\frac{\partial c}{\partial \sigma} = S\sqrt{T}\frac{e^{\left(\frac{-(d_1^2)}{2}\right)}}{\sqrt{2\pi}} > 0 \qquad Vega \qquad \nu \tag{9.29}$$

Although the Black-Scholes model assumes that the underlying stock volatility is constant over time, in reality, volatility can and does shift. Vega provides an estimate for the impact of a volatility shift on the option's value. Vega is also useful in calculating option implied volatilities when the iteration process is made more efficient with the use of the Newton-Raphson method.

A trader should expect that the value of the call would be directly related to the riskless return rate and inversely related to the call exercise price:

$$\frac{\partial c}{\partial e^{r_f}} = TXe^{-r_f T} N(d_2) > 0 \qquad Rho \qquad \rho \tag{9.30}$$

$$\frac{\partial c}{\partial X} = -\,\mathrm{e}^{-r_f T} N(d_2) < 0 \tag{9.31}$$

The option rho can be very useful in economies with high or volatile interest rates, though most traders of "plain vanilla options" do not concern themselves much with call value sensitivities to exercise prices.

Illustration: Greeks Calculations for Calls

In the appendix to Chapter 7, we worked through the Black-Scholes model to value a call. We summarize details as follows:

$$T = 0.5 \quad r_f = 0.10$$

$$X = 80 \quad \sigma^2 = 0.16$$

$$\sigma = 0.4 \quad S_0 = 75$$

$$d_1 = \frac{\ln(75/80) + (0.1 + 0.5 \cdot 0.16)0.5}{0.4 \cdot \sqrt{0.5}} = \frac{\ln(0.9375) + 0.09}{0.2828} = 0.09$$

$$d_2 = d_1 - 0.4 \cdot \sqrt{0.5} = 0.09 - 0.2828 = -0.1928$$

$$N(d_1) = N(0.09) = 0.536 \qquad N(d_2) = N(-0.1928) = 0.42$$

$$c_0 = 75 \cdot 0.536 - (80 \cdot 0.9512) \cdot 0.42 = 7.96$$

Sensitivities (Greeks) for this call option are computed as follows:

$$\frac{\partial c}{\partial S} = N(d_1) = \Delta = 0.536$$

$$\frac{\partial^2 c}{\partial S^2} = \frac{\partial \Delta}{\partial S} = \frac{\partial N(d_1)}{\partial S} = \frac{\mathrm{e}^{\left(\frac{-(d_1^2)}{2}\right)}}{S\sigma\sqrt{T}} = \Gamma = 0.0187$$

$$\frac{\partial c}{\partial T} = -\,r_f X \mathrm{e}^{-r_f T} N(d_2) - S\frac{\sigma}{\sqrt{T}}\frac{\mathrm{e}^{\left(\frac{-(d_1^2)}{2}\right)}}{2\sqrt{2\pi}} = \theta = -11.6522 \qquad \textit{Daily} \quad \textit{Theta} = -0.032$$

$$\frac{\partial c}{\partial \sigma} = S\sqrt{T}\frac{\mathrm{e}^{\left(\frac{-(d_1^2)}{2}\right)}}{\sqrt{2\pi}} = \nu = 21.07$$

$$\frac{\partial c}{\partial e^{r_f}} = TX\mathrm{e}^{-r_f T} N(d_2) = \rho = 16.1156$$

$$\frac{\partial c}{\partial X} = -\,\mathrm{e}^{-r_f T} N(d_2) = -0.4029$$

Sometimes thetas are expressed in daily terms and sometimes vegas and rhos are expressed in terms based on one percentage point increases in standard deviation or interest rates.

Illustration: Hedging With Delta and Gamma Neutral Portfolios

In our numerical example above, we can establish a delta neutral portfolio by purchasing a single call and shorting 0.536 shares of stock. The portfolio delta is the weighted average of its component security deltas, and one call has already been purchased:

$$delta_S \times \#_S + delta_C \times \#_C = delta_{port}$$

$$\#_C = 1$$

The portfolio delta is to be set equal to zero for delta neutrality, and the delta of shares is 1.

$$1 \times \#_S + 0.536 \times \#_C = 0$$

$$\#_C = 1$$

Clearly the solution to this system is $\#_S = -0.536$, implying that delta neutrality requires an offsetting position in 0.536 shares of underlying stock.

Next, suppose instead that we have purchased a single share of underlying stock and that we wish to maintain both delta and gamma neutrality (Δ and Γ equal zero) for our portfolio. These two criteria cannot simultaneously be accomplished with only a single option with our single share. Suppose that we can take positions in a second call option, with the same exercise terms as the first except that the exercise price equals 75 rather than 80. We calculate the delta and gamma of our second option to be 0.62483 and 0.0178, respectively. Stock deltas are 1 and stock gammas are zero. Now, we will seek to solve the following system:

$$\begin{bmatrix} delta_{port} \\ gamma_{port} \\ \#_S \end{bmatrix} = \begin{bmatrix} delta_S & delta_{C1} & delta_{C2} \\ gamma_S & gamma_{C1} & gamma_{C2} \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \#_S \\ \#_{C1} \\ \#_{C2} \end{bmatrix}$$

$$\mathbf{r} \quad = \quad \mathbf{Y} \quad \#$$

$$\begin{bmatrix} delta_{port} \\ gamma_{port} \\ \#_S \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0.53586 & 0.62483 \\ 0 & 0.0187 & 0.0178 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \#_S \\ \#_{C1} \\ \#_{C2} \end{bmatrix}$$

$$\mathbf{r} \quad = \quad \mathbf{Y} \quad \#$$

$$\begin{bmatrix} \#_S \\ \#_{C1} \\ \#_{C2} \end{bmatrix} = \begin{bmatrix} 1 \\ 8.42 \\ -8.82 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ -8.42 & 294.27 & 8.42 \\ 8.82 & -252.37 & 8.42 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

$$\# \qquad \mathbf{Y}^{-1} \qquad \mathbf{r}$$

Thus, for each underlying share of stock purchased, a delta-gamma neutral portfolio requires a long position in 8.42 calls with \$80 exercise prices and a short position in 8.82 calls with \$75 exercise prices. The availability of additional option contracts would increase flexibility with respect to maintaining delta-gamma neutrality. On the other hand, simultaneous maintenance of theta neutrality along with delta-gamma neutrality would require at least one additional different option contract.

Put Sensitivities

If put−call parity holds along with Black-Scholes assumptions given above, the Black-Scholes put value is computed as follows:

$$p_0 = -S_0 N(-d_1) + \frac{X}{e^{r_f T}} N(-d_2) \tag{9.32}$$

$$d_1 = \frac{\ln(S_0/X) + (r_f + 0.5\sigma^2)T}{\sigma\sqrt{T}} \tag{9.24}$$

$$d_2 = d_1 - \sigma\sqrt{T} \tag{9.25}$$

In our numerical example, the put is worth 9.05 if its exercise terms are identical to those of the call in the same example. We can use these put sensitivities to hedge and manage portfolio risk in the same manner that we used the call sensitivities. Put sensitivities formulas and calculations for our example are as follows:

$$\frac{\partial p}{\partial S} = N(d_1) - 1 = \Delta_p = -0.464$$

$$\frac{\partial^2 p}{\partial S^2} = \frac{\partial \Delta_p}{\partial S} = \frac{\partial N(d_1) - 1}{\partial S} = \frac{\frac{e^{\left(\frac{-(d_1^2)}{2}\right)}}{\sqrt{2\pi}}}{S\sigma\sqrt{T}} = \Gamma_p = 0.0187$$

$$\frac{\partial p}{\partial T} = r_f X e^{-r_f T} N(-d_2) - S\frac{\sigma}{\sqrt{T}}\frac{e^{\left(\frac{-(-d_1^2)}{2}\right)}}{2\sqrt{2\Pi}} = \theta_p = -4.110; \quad \textit{Daily Theta} = -0.011$$

$$\frac{\partial p}{\partial \sigma} = S\sqrt{T}\frac{e^{\left(\frac{-(d_1^2)}{2}\right)}}{\sqrt{2\pi}} = \nu_p = 21.06$$

$$\frac{\partial p}{\partial e^{r_f}} = -TXe^{-r_f T} N(-d_2) = \rho_p = -21.93$$

$$\frac{\partial p}{\partial X} = -e^{-r_f T}(N(d_2) - 1) = 0.548$$

9.5 EXCHANGE OPTIONS

In this section, we discuss a variation of the Black-Scholes model provided by Garman and Köhlagen (1983) and Grabbe (1983) for the valuation of currency options. The stock option pricing model is based on the assumption that investors price calls as though they expect the underlying stock to earn the risk free rate of return. However, when we value currency options, we acknowledge that interest rates might differ between the foreign and domestic countries, and that money might be borrowed in one country and loaned in another. Hence, we quote two interest rates—one each for the foreign and domestic currencies. In this scenario, we value currency options as follows:

$$c_0 = s_0 e^{-r(f)T} N(d_1) - \frac{X}{e^{r(d)T}} N(d_2) \tag{9.33}$$

$$d_1 = \frac{\ln\left(\frac{s_0}{X}\right) + \left(r(d) - r(f) + \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}} \tag{9.34}$$

$$d_2 = d_1 - \sigma\sqrt{T} \tag{9.35}$$

where $r(d)$ is the domestic riskless rate (the rate for the currency that will be given up if the option is exercised), $r(f)$ is the foreign riskless rate (the rate for the currency that might be purchased). The spot rate for the currency is s_0. The standard deviation of proportional exchange rates for the foreign currency to be purchased in terms of the domestic currency is σ. The exercise price of the option, X, represents the number of units of the domestic currency to be given up for each unit of the foreign currency. The following variation of put−call parity will apply to the valuation of currency puts:

$$p_0 = c_0 + Xe^{-r(d)T} - s_0 e^{-r(f)T} \tag{9.36}$$

Illustration: FX Option Valuation

Consider, for example, a currency option series on Swiss francs (CHF) denominated in U.S. dollars. Suppose that a six-month (0.5 years) call option is available on francs with an exercise price equal \$0.5556. The domestic and foreign riskless rates are 0.1. The current exchange rate is \$0.5556/CHF and the standard deviation associated with the exchange rate is 0.4. What is the value of this currency call?

To answer this question, we first calculate d_1:

$$d_1 = \frac{\ln\left(\frac{0.5556}{0.5556}\right) + \left(0.1 - 0.1 + \frac{0.4^2}{2}\right) \cdot 0.5}{0.4\sqrt{0.5}} = \frac{0.04}{0.2828} = 0.1414$$

Next we calculate d_2:

$$d_2 = d_1 - \sigma\sqrt{T} = 0.1414 - 0.2828 = -0.1414$$

Next, find cumulative normal density functions (z-values) for d_1 and d_2:

$$N(d_1) = N(0.1414) = 0.556$$
$$N(d_2) = N(-0.1414) = 0.443$$

Finally, we value the call as follows:

$$c_0 = (\$0.5556 \times 0.556) - [\$0.5556 \times 0.9512] \times 0.443 = \$0.05943$$

We can evaluate a European put for CHF using Equation (9.36) as follows:

$$p_0 = c_0 + Xe^{-r(d)T} - s_0e^{-r(f)T} = 0.05943 + 0.5556e^{-0.1 \times 0.5} - 0.5556e^{-0.1 \times 0.5} = 0.05943$$

Consider the next example where four call options are traded on CHF (Swiss Francs) with exercise prices X expressed in U.S. dollars. One U.S. dollar is currently worth 2 Swiss francs ($s_0 = 2$). We wish to evaluate calls for each of the following European currency option series depicted in Table 9.2.

We will work through Equations (9.33) through (9.35) to value the calls. Final solutions are given in the bottom line of Table 9.3.

We use a currency-specific variation of the put–call parity relation (Equation 9.36) to value the puts:

$$p_0 = c_0 + Xe^{-r(d)T} - s_0e^{-r(d)T}$$

TABLE 9.2 Currency Option Illustration

Option 1	Option 2	Option 3	Option 4
$T = 1$	$T = 1$	$T = 0.5$	$T = 2$
$r(f) = 0.1$	$r(f) = 0.1$	$r(f) = 0.1$	$r(f) = 0.1$
$r(d) = 0.05$	$r(d) = 0.05$	$r(d) = 0.05$	$r(d) = 0.05$
$s_0 = 2$	$s_0 = 2$	$s_0 = 2$	$s_0 = 2$
$\sigma = 0.3$	$\sigma = 0.3$	$\sigma = 0.3$	$\sigma = 0.3$
$X = 1.8$	$X = 2$	$X = 2.2$	$X = 2.2$

TABLE 9.3 Currency Option Values

	Option 1	Option 2	Option 3	Option 4
d_1	0.3345	−0.0167	−0.4611	−0.2482
d_2	0.0345	−0.3167	−0.6732	−0.6725
$N(d_1)$	0.6310	0.4934	0.3224	0.4020
$N(d_2)$	0.5138	0.3757	0.2504	0.2506
Call	0.2622	0.1780	0.0760	0.1593

Thus, put options with terms identical to those of the calls in Tables 9.2 and 9.3 have the following values:

	Option 1	Option 2	Option 3	Option 4
Put	0.1648	0.2707	0.3192	0.5125

9.6 HEDGING EXCHANGE EXPOSURE WITH CURRENCY OPTIONS

Based on the example from Chapter 8, Hedging Exchange Exposure section, suppose that the Dayton Company of America expects to receive a payoff of £1,000,000 in three months, and intends to convert its cash flows to dollars. Continue to assume relevant data as follows:

Spot exchange rate: $1.7640/£
Three-month forward exchange rate: $1.7540/£
U.K. borrowing interest rate: 10.0%
U.S. borrowing interest rate: 8.0%
U.K. lending interest rate: 8.0%
U.S. lending interest rate: 6.0%

Also assume that there exist call and put options and forward contracts with the following terms, and that their premiums might not reflect formula values:

Term to options expiration: 3 months
Exercise price: $1.75/£
Put premium: $0.025/£
Call premium: $0.065/£
Brokerage cost per options contract on £31,250: $50

Our problem is to evaluate methods of managing the transaction risk associated with this extension of credit and the implications of each.

We will consider two options-based hedging strategies here. The first is the put hedge (partial hedge) strategy which involves the purchase of a put on pounds, enabling the firm to protect itself against devaluation of pounds. If the value of pounds increases, the firm realizes a greater profit. However, the firm must pay the full cost of the put. With the conversion or call and put hedge strategy, the proceeds from the sale of a call are used to offset the purchase price of the put. This strategy acts as a collar, locking in the value of pounds at the originations of the options contracts. Hence, the firm would not benefit from any appreciation in the value of the pound.

First, we consider the put hedge strategy. We will purchase three month put options on £1,000,000 with an exercise price of $1.75/£ with a total premium of $25,000. Time-zero brokerage costs total $1600 (32 contracts at $50 per contract). Thus the total time-zero cash outlay is $26,600. Forgone interest on the sum of the premium and brokerage costs totals $399. Expressed in terms of future value, the total cash outlay is $26,999. The result of this strategy is that the firm receives one of the following in three months:

1. An unlimited maximum less the $26,999 premium, forgone interest and brokerage fees. The dollar value of this strategy increases as the value of the dollar drops against the pound. Since cash flows are not certain, this hedge is considered partial.
2. A minimum of $1,750,000 less $26,999 for a net of $1,723,001. This minimum value to be received might be unacceptably low; however, there is upside cash flow potential.

Alternatively, the firm can employ the conversion or the call and put hedge. This strategy involves the combination of calls and puts, such that total risk can be eliminated. Consider the writing of a call with an exercise price of $1.75 expiring in three months along with the purchase of a put with the same terms. The time-zero net cash flows are summarized as follows:

Put premium − $25,000	Call premium + $65,000
Put brokerage fee − $1600	Call brokerage fee − $1600
Net time-zero cash flows + $36,800	

The result of this conversion is that the interest earned on the net time-zero outlay is $552. If the three-month exchange rate is less than $1.75/£, the exchange rate of $1.75/£ is locked in by the put. If the exchange rate exceeds $1.75/£, the obligation incurred by the short position in the call is activated. Thus, the firm's exchange rate of $1.75/£ is locked in no matter what the market exchange rate is. The cash flows in three months are summarized as follows:

Put cash flows (£1,000,000 × MAX[1.75 − S_1,0])
Call cash flows (£1,000,000 × MIN[1.75 − S_1,0])

Total of option transactions:

$$\begin{aligned} £1,000,000 \times (1.75 - S_1) &= \$1,750,000 - (£1,000,000 \times S_1) \\ \text{Exchange of currency} &= (£1,000,000 \times S_1) \\ \text{Time zero cash flows} &= \$36,800 \\ \text{Interest on time zero flows} &= \$552 \\ \text{Total time one cash flows} &= \$1,787,352 \end{aligned}$$

This cash flow of $1,787,352 is assured in the absence of default risk.

Additional Reading

Elton, Gruber, Brown, and Goetzmann (2010), Chapters 21 to 24, offer good coverage of fixed income, options and futures arbitrage and hedging. Grabbe (1996) discusses currency option pricing and hedging. Jarrow and Rudd (1983), while dated, provide a very useful overview of options pricing, hedging and arbitrage. The industry standard reading for options and derivatives is Hull (2011), going far beyond the scope and level of rigor of material in this chapter. Neftci (2000) provides excellent background reading in the mathematics of financial derivatives, again, going far beyond the scope of this chapter.

References

Cox, J., Ross, S., & Rubinstein, M. (1979). Option pricing: A simplified approach. *Journal of Financial Economics*, *7*, 229–263.
Elton, E., Gruber, M., Brown, S., & Goetzmann, W. (2010). *Modern portfolio theory and investment analysis* (8th ed.). New York: Wiley.
Garman, M., & Köhlagen, S. (1983). Foreign currency option values. *Journal of International Money and Finance*, *2*, 231–237.
Grabbe, J. O. (1983). The pricing of call and put options on foreign exchange. *Journal of International Money and Finance*, *2*, 239–253.
Grabbe, J. O. (1996). *International financial markets* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.
Hull, J. C. (2011). *Options, futures and other derivatives* (8th ed.). Englewood Cliffs, NJ: Prentice Hall.
Jarrow, R., & Rudd, A. (1983). *Option pricing*. Homewood, IL: Dow Jones-Irwin.
Jarrow, R. A., & Turnbull, S. M. (1999). *Derivative securities* (2nd ed.). Cincinnati: South-Western College Publishers.
Neftci, S. N. (2000). *An introduction to the mathematics of financial derivatives* (2nd ed.). San Diego: Academic Press.

9.7 EXERCISES

1. Some observers of derivatives markets agree that publicly traded corporations can use derivatives to mitigate operating risks but that shareholders would be better off if managers simply allowed shareholders to use their own portfolio management and hedging techniques to do so themselves. In fact, some critics argue that corporate managers are too concerned about mitigating risks and that they do so to protect their own jobs and preserve their own income and stock value.
   a. Why might shareholders be in a better position to mitigate and hedge operating risks of corporations whose shares they own?
   b. Why might shareholders be better off if corporate managers continue to hedge operating and financial risks?
2. Consider a one-time-period, two potential outcome framework where there exists Company Q stock currently selling for $50 per share and a riskless $100 face value T-Bill currently selling for $90. Suppose Company Q faces uncertainty, such that it will pay its owner either $30 or $70 in one year. Further assume that a call with an exercise price of $55 exists on one share of Q stock.
   a. What are the two potential values the call might have at its expiration?
   b. What is the riskless rate of return for this example? Remember, the Treasury bill pays $100 and currently sells for $90.
   c. What is the hedge ratio for this call option?
   d. What is the current value of this option?
   e. What is the value of a put with the same exercise terms as the call?
3. Rollins Company stock currently sells for $12 per share and is expected to be worth either $10 or $16 in one year. The current riskless return rate is 0.125. What would be the value of a one-year call with an exercise price of $8?
4. A stock currently selling for $50 has an annual returns variance equal to 0.36. The riskless return rate equals 0.08 per year. Under the binomial framework, what would be the value of nine-month (0.75 year) European calls and European puts with striking prices equal to $80 if the number of tree steps (n) were

a. two?
b. three?
c. eight?

5. Ibis Company stock is currently selling for $50 per share and has a multiplicative upward movement equal to 1.2776 and a multiplicative downward movement equal to 0.7828. What is the value of a nine-month (0.75 year) European call and a European put with striking prices equal to $60 if the number of tree steps were two? Assume a riskless return rate equal to 0.081.

6. Stanton Company stock is trading for 50 in a two-time-period environment, with each relevant time period lasting 6 months. The stock might increase by exactly 20% in just one period or perhaps in both periods. Of course, the stock might not increase in either period. If the stock price does not increase in a given period, it will decline by 16.67 percent in that particular period. One-year options with an exercise price equal to 60 are trading on this stock. The annual riskless rate of return equals 0.
a. What is the value of a put in this environment?
b. What is the probability (risk-neutral probability) implied in this framework that the Stanton Company stock price will exceed 40 when options expire in one year?

7. Kestrel Company stock is currently selling for $40 per share. Its historical standard annual deviation of returns is 0.5 in this single-period environment. The one-year Treasury bill rate is currently 5%. Assume that all of the standard Black-Scholes option pricing model assumptions hold.
a. What is the value of a put on this stock if it has an exercise price of $35 and expires in one year?
b. What is the implied probability that the value of the stock will be less than $30 in one year?

8. Evaluate each of the European options in the series on ABC Company stock. Prices for each of the options are listed in the table. Determine whether each of the options in the series should be purchased or sold at the given market prices. The current market price of ABC stock is 120, the August options expire in nine days, September options in 44 days, and October options in 71 days. The stock return standard deviations prior to expirations are projected to be 0.20 prior to August, 0.25 prior to September, and 0.20 prior to October. The Treasury bill rate is projected to be 0.06 for each of the three periods prior to expiration. Do not forget to convert the number of days given to fractions of 365-day years.

Calls

X	Aug	Sep	Oct	
110	9.500	10.500	11.625	$\sigma = 0.20$ for Aug
115	4.625	7.000	8.125	$\sigma = 0.25$ for Sep
120	1.250	3.875	5.250	$\sigma = 0.20$ for Oct
125	0.250	2.125	3.125	$r = 0.06$
130	0.031	0.750	1.625	$S = 120$

Puts

X	Aug	Sep	Oct
110	0.031	0.750	1.500
115	0.375	1.750	2.750
120	1.625	6.750	4.500
125	5.625	6.750	7.875
130	10.625	10.750	11.625

Prices for five calls and five puts are given in the left columns. Expiration dates are given in column headings and current market prices are given in the table interiors.

9. What is the gamma of a long position in a single futures contract? Why?

10. Tanker Company stock currently sells for 30 per share and has an anticipated volatility (annual return standard deviation) equal to 0.6. Three-month (0.25 year) call options are available on this stock with an exercise price equal to 25. The current riskless return rate equals 0.05.
- **a.** Calculate the call's value.
- **b.** Calculate the call's delta, gamma, theta, vega, and rho.
- **c.** What is the Black-Scholes implied probability that the stock price will exceed 25 three months from now?
- **d.** What is the value of a put with the same exercise terms as the call?
- **e.** What are the Greeks for the put in part d?

11. Tanker Company stock, which currently sells for 30 per share and has an anticipated volatility (annual return standard deviation) equal to 0.6, also has a three-month (0.25 year) call option with an exercise price equal to 30. The current riskless return rate equals 0.05.
- **a.** Calculate this second call's value.
- **b.** Calculate this call's delta, gamma, theta, vega, and rho.
- **c.** What is the Black-Scholes implied probability that the stock price will exceed 30 three months from now?
- **d.** What positions in these calls and the calls from Question 10 should be taken for each long share of stock to maintain a delta-gamma neutral portfolio?
- **e.** What is the theta of the portfolio in part d?

12. A currency option series on Australian dollars (AUD) denominated in U.S. dollars. Suppose that a two-year call option is available on AUD with an exercise price equal USD0.65. The U.S. interest rate is 0.04 and the Australian rate is 0.06. The current exchange rate is USD0.7/AUD and the standard deviation associated with the exchange rate is 0.3. What is the value of this currency call? What is the value of a put with the same exercise terms?

13. Suppose that right now, the market price of one Chinese yuan (CNY) is USD0.15; that is, one can purchase or sell CNY1.00 for USD0.15. The proportional standard deviation for CNY in terms of USD is 20%. The Chinese riskless return rate over a two-year period is projected to be 8%; U.S. rates over the same period are 2%. What is the dollar

value of a two-year European put on a single yuan if the exercise price of the put is USD0.20?

14. Consider an extension of the Exercise 21 from Chapter 8 concerned with the management of transactions exposure. The Smedley Company sold products to a Japanese client for ¥15,000,000. Payment is due six months later. Relevant data are as follows:

Spot exchange rate: ¥105/$
Japanese borrowing interest rate: 9.0%
U.S. borrowing interest rate: 7.0%
Japanese lending interest rate: 7.0%
U.S. lending interest rate: 5.0%
There exist currency call and put options with the following terms:
Size of options contracts: ¥1,000,000
Term to expiration of options contracts: six months
Exercise price of put and call: $0.009/¥
Put premium: $0.00001/¥
Call premium: $0.0001/¥
Brokerage cost per options contract: $50

Discuss the implications associated with each of the options-based methods for managing the transactions exposure risk associated with this extension of credit.

APPENDIX

9.A

9.A.1 THE BINOMIAL MODEL: ADDITIONAL CONSIDERATIONS

A Computationally More Efficient Version of the Binomial Model

Here, we simply discuss a computationally more efficient version of the n-time-period binomial model Equation (9.17):

$$c_0 = \frac{\sum_{j=0}^{n} \frac{n!}{j!(n-j)!} p^j (1-p)^{n-j} \text{MAX}\left[0, u^j d^{n-j} S_0 - X\right]}{(1+r_f)^n} \tag{9.17}$$

The number of computational steps required to solve Equation (9.17) is reduced if we eliminate from consideration all of those outcomes where the option's expiration date price is zero. Thus, the smallest non-negative integer for j where $S_n > X$ is the smallest integer that exceeds the following from Equation (9.19):

$$a = \text{MAX}\left[\frac{\ln\left(\frac{X}{S_0 d^n}\right)}{\ln\left(\frac{u}{d}\right)}, 0\right] \tag{9.19}$$

This result is obtained by first determining the minimum number of price increases j needed for S_n to exceed X:

$$S_n = u^j d^{n-j} S_0 > X \tag{9.A.1.1}$$

We then solve this inequality for the minimum non-negative integer value for j such that $u_j d^{n-j} S_0 > X$. Take logs of both sides to obtain:

$$\begin{aligned} j \cdot \ln(u) + n \cdot \ln(d) - j \cdot \ln(d) + \ln(S_0) &> \ln X \\ j \cdot \ln\left(\frac{u}{d}\right) &> \ln(X) - n \cdot \ln(d) - \ln(S_0) \end{aligned} \tag{9.A.1.2}$$

Next divide both sides by $ln(u/d)$ and simplify to obtain Equation (9.19), where a is the smallest non-negative integer for j for which $S_n > X$. Finally, we substitute Equation (9.19) into Equation (9.7) to obtain *the binomial option pricing model*:

$$c_0 = \frac{\sum_{j=a}^{n} \frac{n!}{j!(n-j)!} p^j (1-p)^{n-j} \left[u^j d^{n-j} S_o - X\right]}{(1+r_f)^n} \tag{9.18}$$

Or

$$c_0 = S_0\left[\sum_{j=a}^{n}\frac{n!}{j!(n-j)!}\frac{(pu)^j[(1-p)d]^{n-j}}{(1+r_f)^n}\right] - \frac{X}{(1+r_f)^n}\left[\sum_{j=a}^{n}\frac{n!}{j!(n-j)!}p^j(1-p)^{n-j}\right] \tag{9.18a}$$

In short-hand form:

$$c_0 = S_0B_1[a,n,p'] - X(1+r_f)^{-n}B_2[a,n,p] \tag{9.18b}$$

where $p' = pu/(1+r_f)$. There are two more points regarding Equation (9.18b). First, assuming that investors behave as though they are risk neutral, $B_2[a, n, p]$ can be interpreted as the probability that the stock price will be sufficiently high at the expiration date of the option to warrant its exercise. Second, $B_1[a, n, p']$ can be interpreted as a hedge ratio, although it must be updated at every period.

Obtaining Multiplicative Upward and Downward Movement Values

One apparent difficulty in applying the binomial model as it is presented above is in obtaining estimates for u and d that are required for p; all other inputs are normally quite easily obtained. There are several methods that are used to obtain parameters for the binomial method from the actual security returns generating process. For example, following Cox et al. (1979), we can begin the process of estimating the mean and variance to be used in the binomial distribution by first approximating the mean and variance for the binomial process from the historical Wiener process as follows:

$$E[S_T] = S_0e^{\mu T+\frac{1}{2}\sigma^2T} \approx puS_0 + (1-p)dS_0 \tag{9.A.1.3}$$

$$\sigma_s^2T = S_0^2(e^{\sigma^2T}-1)(e^{2\mu T+\sigma^2T}) \approx [pu^2S_0^2 + (1-p)d^2S_0^2] - [puS_0+(1-p)dS_0]^2 \tag{9.A.1.4}$$

Approximation 2 approaches equality as T approaches zero. Scaling S_0 to one such that we work with returns rather than actual security prices, the following can be used for returns variance of a binomial process:

$$\sigma^2 = [pu^2 + (1-p)d^2] - [pu+(1-p)d]^2 \tag{9.A.1.5}$$

We will rewrite Equation (9.A.1.5) as follows[11]:

$$\sigma^2 = p(1-p)(u-d)^2 \tag{9.A.1.6}$$

[11]The following are algebraic steps to obtain Equation (9.A.1.6) from Equation (9.A.1.5):

$\sigma^2 = \lfloor pu^2 + (1-p)d^2\rfloor - [pu+(1-p)d]^2 = \lfloor pu^2 + d^2 - pd^2\rfloor - \lfloor p^2u^2 + (1-p)^2d^2 + 2pu(1-p)d\rfloor$

$\sigma^2 = \lfloor pu^2 + d^2 - pd^2\rfloor - \lfloor p^2u^2 + d^2 + p^2d^2 - 2pd^2 + 2pud - 2p^2ud\rfloor$

$\sigma^2 = |pu^2| - |p^2u^2 + p^2d^2 - pd^2 + 2pud - 2p^2ud|$

$\sigma^2 = p(u^2 - pu^2 - pd^2 + d^2 - 2ud + 2pud) = p(1-p)(u^2 + d^2 - 2ud)$

Assume that the binomial process will lead to expected return for a security equaling the riskless rate:

$$e^{\mu T+\frac{1}{2}\sigma^2 T} = e^{r_f} \approx pu + (1-p)d \tag{9.A.1.7}$$

Solving Equation (9.A.1.7) for p enables us to estimate probabilities of an uptick p and downtick $(1-p)$ as:

$$p = \frac{e^{r_f} - d}{u - d} \quad (1-p) = \frac{u - e^{r_f}}{u - d} \tag{9.A.1.8}$$

If we define d as $1/u$ such that $ud = 1$, we can rewrite Equation (9.A.1.6), the variance of returns, as follows:

$$\sigma^2 = p(1-p)(u-d)^2 = p(1-p)(e^{\delta} - e^{-\delta})^2 \tag{9.A.1.9}$$

Thus, we have simply substituted some constant e^{δ} for u. Substituting σ for δ will provide a good approximation for variance (improving as the number of jumps in the binomial process, n approaches infinity):

$$\sigma^2 = p(1-p)(e^{\delta} - e^{-\delta})^2 = p(1-p)(e^{\sigma} - e^{-\sigma})^2 \tag{9.A.1.10}$$

Thus, we can use the following to estimate u and d in the binomial approximation to the Wiener process:

$$u = e^{\sigma} \quad d = \frac{1}{u} \tag{9.A.1.11}$$

or, if n and T differ from 1,

$$u = e^{\sigma\sqrt{\frac{T}{n}}} \quad d = \frac{1}{u} \tag{9.A.1.12}$$

Suppose, for example, that for a particular Wiener process, $\sigma = 0.30$ and $r_f = 0.05$. Using Equations (9.A.1.7) and (9.A.1.10), we estimate p, u, and d for a single time period binomial process as follows:

$$u = e^{0.3} = 1.3498588$$

$$d = \frac{1}{u} = 0.7408182$$

We can verify our estimates with Equations (9.A.1.6) and (9.A.1.9) as follows:

$$e^{0.005+\frac{1}{2}\times 0.3^2} = e^{0.05}$$

$$= 0.5097409 \times 1.3498588 + (1 - 0.5097409) \times 0.7408182 = 1.0512$$

$$\sigma^2 = p(1-p)(e^{\delta} - e^{-\delta}) \approx 0.5097409 \times (1 - 0.5097409)(e^{0.3} - e^{-0.3})^2 = 0.0926974$$

As discussed above, there are several procedures for getting parameters σ, u, d, and p for the binomial distribution. This procedure is probably the most commonly used, in part, because it provides a relatively straightforward method for estimating option Greeks.

The primary difficulty with the one presented above is that it may result in negative probabilities. An additional methodology for estimating binomial distribution parameters is given in Jarrow and Turnbull (1999: 136–138).

9.A.2 DERIVING THE BLACK-SCHOLES MODEL

For this derivation, we shall assume that all standard Black-Scholes assumptions (see The Greeks and Hedging in a Black-Scholes Environment section) hold and that investors behave as though they are risk neutral. That is, investors price options as though they are risk neutral because they can always construct riskless hedges comprising options and their underlying securities, and that the returns of these hedged portfolios must be the riskless rate. The law of one price dictates that the current value of a call c on stock can be found from constructing a hedge portfolio:

$$c = \#sS + \#bB$$

S is the current value of its underlying stock and B is the current value of a riskless Treasury instrument. Let $\#s$ be the number of shares of stock to purchase and $\#b$ be the number of Treasury instruments to short in order to replicate the cash flow structure of the call. Similarly, we can replicate the cash flow structure of a single bond as follows:

$$B = \#sS + c \tag{9.A.2.1}$$

Let $V_h = B$ represent the value of a perfectly hedged portfolio. We can rewrite the above equation in terms of VH as follows:

$$V_h = c + \#sS \tag{9.A.2.2}$$

Since the hedge is riskless, its return should equal the riskless rate:

$$dV_h = r_f V_h dt = dc + \#sdS \tag{9.A.2.3}$$

The hedge requires that we short sell $\#s$ shares of stock for each call that we purchase. The sensitivity of the call price to the stock price is $\partial c/\partial S$. Thus, the hedge will require that we short $\partial c/\partial S$ shares for each purchased call. We write the value of the hedge portfolio and rewrite its differential equation as follows:

$$V_h = c - \frac{\partial c}{\partial S} S \tag{9.A.2.4}$$

$$r_f V_h dt = dc - \frac{\partial c}{\partial S} dS \tag{9.A.2.5}$$

We rearrange the above differential equation by substituting for V_h and solving for dc:

$$dc = r_f \left[c - \frac{\partial c}{\partial S} S \right] dt + \frac{\partial c}{\partial S} dS \tag{9.A.2.6}$$

which is rewritten as follows:

$$dc = r_f c dt - r_f \frac{\partial c}{\partial S} S dt + \frac{\partial c}{\partial S} dS \tag{9.A.2.7}$$

We shall assume that the instantaneous price change for the stock follows an Itô process:

$$dS = \mu S dt + \sigma S dz \tag{9.A.2.8}$$

which requires us to use Itô's lemma to solve for dc. Substituting this Wiener process for dS into the preceding equation, we obtain the following:

$$dc = r_f c dt - r_f \frac{\partial c}{\partial S} S dt + \frac{\partial c}{\partial S} (\mu S dt + \sigma S dz) \tag{9.A.2.9}$$

We will rewrite Equation (9.A.2.9) separating out the term in the parentheses as follows:

$$dc = r_f c dt - r_f \frac{\partial c}{\partial S} S dt + \frac{\partial c}{\partial S} \mu S dt + \frac{\partial c}{\partial S} \sigma S dz \tag{9.A.2.10}$$

$$dc = \mu S \frac{\partial c}{\partial S} dt + \frac{\partial c}{\partial t} dt + \frac{1}{2} S^2 \sigma^2 \frac{\partial^2 c}{\partial S^2} dt + \sigma S \frac{\partial c}{\partial S} dz \tag{9.A.2.11}$$

Note that the first and fourth terms in Equation (9.A.2.11) are identical to the third and fourth terms in Equation (9.A.2.10). Furthermore, Equations (9.A.2.10) and (9.A.2.11) solve for dc. This means that the second and third terms in Equation (9.A.2.11) are identical to the first and second terms in Equation (9.A.2.10):

$$r_f c dt - r_f \frac{\partial c}{\partial S} S dt = \frac{\partial c}{\partial t} dt + \frac{1}{2} S^2 \sigma^2 \frac{\partial^2 c}{\partial S^2} dt \tag{9.A.2.12}$$

which, after dividing by dt, is rewritten:

$$\frac{\partial c}{\partial t} = r_f c - r_f \frac{\partial c}{\partial S} S_t - \frac{1}{2} S^2 \sigma^2 \frac{\partial^2 c}{\partial S^2} \tag{9.A.2.13}$$

This is the Black-Scholes differential equation. Its particular solution, subject to the *boundary condition* $c_T = \text{MAX}[0, S_T - X]$, is given by Equations (9.A.2.14) through (9.A.2.16):

$$c_0 = S_0 N(d_1) - \frac{X}{e^{r_f T}} N(d_2) \tag{9.A.2.14}$$

$$d_1 = \frac{\ln\left(\frac{S_0}{X}\right) + \left(r_f + \frac{1}{2}\sigma^2\right) T}{\sigma \sqrt{T}} \tag{9.A.2.15}$$

$$d_2 = d_1 - \sigma \sqrt{T} \tag{9.A.2.16}$$

where $N(d^*)$ is the cumulative normal distribution function for (d^*). This function is frequently referred to in a statistics setting as the z-value for (d^*). This solution to the Black-Scholes differential equation can be verified by finding the derivative of c_0 in the Black-Scholes model with respect to t. From a computational perspective, one would first work through Equation (9.A.2.15) and then Equation (9.A.2.16) before valuing the call with Equation (9.A.2.14).

CHAPTER

10

Evaluating Trading Strategies and Performance

10.1 EVALUATING INVESTMENT PORTFOLIO PERFORMANCE

A goal of every investor is to find an investment system that generates large profits. An *investment system* is simply a security selection and combination strategy. Huge varieties of systems are used by investors, ranging from making purchase decisions based on moving averages to selling securities based on sunspot cycle theories. Many of these systems have made millions of dollars for investors. It should be noted that while perhaps each of these systems have worked well at some point in the past for some investors, or at least some investors have been lucky using them in the past, they might not perform well in the future. With thousands of investment systems in practice at any point in time, at least one might be expected to perform well over a period of time in any market, inasmuch as at least one out of thousands of randomly selected portfolios might be expected to perform well.

Among the simplest of investment systems is the simple buy and hold into a diversified portfolio strategy. A large randomly selected portfolio or index portfolio might be used as a benchmark portfolio for comparison of any other investment system. Proper evaluation of an investment system does require consideration of a number of factors and statistical concerns. Perhaps, the most important and difficult of these to account for is risk. One should note that higher-risk portfolios would be expected to generate higher returns over extended periods of time. Furthermore, the evaluator might also wish to account for investor risk preferences or objectives when evaluating an investment system.

One problem that we encounter when evaluating trading strategies and performance is separating the portfolio selection from the trading decisions. Many traders make both decisions simultaneously. In this section, we will focus first on the portfolio decision performance aspect of the trading decision, leaving trade execution aspects to later sections. The first step in the evaluation of portfolio decision performance is to measure the increase in asset or portfolio value relative to the initial asset value during a given time frame. These returns can then be adjusted by or compared to a relevant benchmark. In addition, the trader will want to evaluate the trading component of

Financial Trading and Investing
DOI: https://doi.org/10.1016/B978-0-323-90955-6.00010-0

© 2023 Elsevier Inc. All rights reserved.

performance. This trading decision is often more difficult to evaluate than the overall portfolio performance.

Computing Net Asset Value and Returns

Again, investment portfolio performance can be rather difficult to assess. First, standardization of performance measures and selection of benchmarks are crucial for evaluations and portfolio comparisons. In the case where we evaluate a hypothetical mutual fund, we begin by valuing its net assets at the end of a given trading period, typically at the end of a day, with NAV_t computed as follows:

$$NAV_t = \frac{\text{Market value of assets minus liabilities at the end of time } t}{\text{\# Shares outstanding at time } t} \tag{10.1}$$

Obviously, NAV depends more on the initial investment amount than on portfolio performance. Thus, it is useful to evaluate returns, which are NAV levels measured relative to initial investment amounts or prior NAV levels. If no capital is added or withdrawn from the fund between times $t-1$ and t, the fund's return r_t for time t is computed from the change in its NAV during the period t:

$$r_t = \frac{NAV_t - NAV_{t-1}}{NAV_{t-1}} \tag{10.2}$$

Equation (10.1) provides a one-year return for the portfolio. Now, let's consider the performance of a fund held for n years. It will be convenient to express returns on an annual basis, as is the convention for most institutions. The measurement of a geometric mean annual return r_g for a portfolio held for n years is fairly straightforward when there are no funds added to or withdrawn from the portfolio:

$$\bar{r}_g = \sqrt[n]{\frac{NAV_n}{NAV_0}} - 1 \tag{10.3}$$

However, matters can become more complicated when the fund's investment base is changing. For example, fund investors might deposit and withdraw funds each day. When there is a cash flow CF_t moving in (a deposit to the fund would be represented by a positive value) or out of the portfolio (such as a withdrawal or redemption) at the beginning of each time period t for n periods, one might measure the n-period return with a variation of the internal rate of return r, solving the following for r:

$$NAV_n = NAV_0(1+r)^n - \sum_{t=1}^{n} CF_t(1+r)^{n-t} \tag{10.4}$$

A return computed in this manner is called a *dollar-weighted return*. For example, consider the case where a given cash flow CF_t was positive because the portfolio received a deposit from an investor. This increases the investment base of the portfolio, reducing the portfolio's return if no compensating increase in the terminal asset value NAV_n is realized. This dollar-weighted return accounts for the addition (or subtraction) of cash flows to (or from)

the fund. However, it does not fully account for compounding of returns over time. Bear in mind that computing this dollar-weighted return can be time consuming because substitution is usually required to obtain r and there can be instances when more than one r value will solve this equation.

A second method for computing a return, the *time-weighted average return*, requires periodic return computations each time a change to the investment base occurs:

$$\overline{r}_{g,p} = \sqrt[n]{\prod_{t=1}^{n}(1 + r_t)} - 1 \tag{10.5}$$

where r_t, adjusted for dividends, is the fund return for a single period t and CF_t represents capital moving in (positive) or out of (negative) the fund at the start of each period t for n periods:

$$r_t = \frac{NAV_t - NAV_{t-1} + DIV_t - CF_t}{NAV_{t-1}} \tag{10.6}$$

Generally, the time-weighted return will be a more meaningful measure because the returns on funds initially deposited into the portfolio will not be affected by future deposits or withdrawals.

Net Asset Value and Returns Illustration

In Table 10.1, we illustrate NAVs and returns for a hypothetical Fund A for each of 10 days starting on January 3, 2017. Note that January 4 is a Friday, and that the fund will not trade over the weekend of January 5 and 6. A three-day return is calculated for January 9; the returns for January 4, 5, and 6 will be −3.92%/3 or −1.31% each day.

On January 2, the fund initiates activities with a contribution of $10,000,000 by investors. No return is computed. On January 3, the fund realizes $200,000 in trading profits, increasing its total assets to $10,200,000, yielding a return of 2.00% for the day. On

TABLE 10.1 Net Asset Value and Return Calculations

Date	Investor Contributions	Investor Withdrawals	Fund Dividends	Total Assets	Number of Shares	*NAVt*	r_t
1/2/2017	10,000,000	0	0	10,000,000	1,000,000	10.00	N/A
1/3/2017	0	0	0	10,200,000	1,000,000	10.20	2.00%
1/4/2017	0	0	0	9,800,000	1,000,000	9.80	−3.92%
1/7/2017	0	0	0	10,000,000	1,000,000	10.00	2.04%
1/8/2017	0	0	200,000	10,000,000	1,000,000	10.00	2.00%
1/9/2017	0	200,000	0	10,000,000	980,392	10.20	2.00%
1/10/2017	200,000	0	0	10,500,000	999,429	10.51	3.00%
1/11/2017	300,000	200,000	0	10,400,000	1,009,132	10.31	−1.90%

January 4, the fund incurred $400,000 in trading losses, ending the day with $9,800,000 in total assets, $9.8 in NAV and a return of −3.92% for the day, or −1.31% for Friday, Saturday, and Sunday. Trading profits for January 7 were $200,000; the return was 2%. On January 8, the fund realized $200,000 in trading profits, but paid a dividend of $200,000. NAV remained constant at $10.00 because the trading profits exactly offset the dividends. The return for the day was 2%. On January 9, the fund realized trading profits of $200,000, and investors redeemed 19,608 shares at a value of $10.20 for a total of $200,000. This withdrawal did not affect NAV or returns. On January 10, investors purchased 19,037 shares for $10.51 each, totaling $200,000, which combined with trading profits of $300,000 increasing total assets to $10,500,000. The investor contributions to the fund did not affect its 3% return for the day. On January 11, investor share purchases exceeded share redemptions by 9703. Trading losses for the day totaled $200,000, producing a return of −1.9%. The time-weighted average return for the fund is computed as follows:

$$\bar{r}_{g,p} = \sqrt[n]{\prod_{t=1}^{n}(1+r_t)} - 1 = \sqrt[9]{(1.02)(0.9969)^3(1.0204)(1.02)(1.02)(1.03)(0.981)} - 1 = 0.0056$$

One minor note: The average daily weekend return of −1.31% was computed as an arithmetic average; had it been computed as a geometric mean, it would have been −1.325%.

Portfolio Benchmarking

Higher returns are generally associated with higher risk. Proper risk-adjusted benchmarks for comparison need to be established when evaluating portfolio returns. Portfolio risk can be measured by portfolio standard deviation σ_p. This portfolio standard deviation measures the risk of the portfolio as a function of the risk levels of individual securities, the investment proportions in the individual securities, and the level of diversification of the portfolio. Portfolio risk can also be measured by the sensitivity of the portfolio's return to returns on the market portfolio, captured by the portfolio's beta, β_p, a measure of undiversifiable risk as per the Capital Asset Pricing Model (CAPM).[1] The following are three well-known risk adjusted measures for portfolio performance evaluation (see Elton, Gruber, Brown, & Goetzman, 2010):

$$S_p = \frac{r_p - r_f}{\sigma_p} \qquad T_p = \frac{r_p - r_f}{\beta_p}$$

Sharpe ratio *Treynor ratio*

$$\textit{Jensen measure: } J_p = [r_p - r_f] - [\beta_p(r_m - r_f)]$$

The *Sharpe and Treynor ratios* calculate the portfolio's reward to risk ratio, as a function of overall portfolio risk and as a function of portfolio sensitivity to the market. The *Jensen*

[1]A security beta, β_i or portfolio beta, β_p, can be computed as the slope term from a simple OLS regression of risk premiums on the security $(r_i - r_f)$ or portfolio $(r_p - r_f)$ against risk premiums of the market $(r_m - r_f)$. See text Appendix Section A.1 for a brief review of the simple OLS.

measure (Jensen, 1968) is often referred to as the portfolio alpha. Alpha measures the excess return of a portfolio beyond the time value of money and risk. A host of other measures have been described in the public literature, including the Sortino coefficient (Sortino, 1996) and the M-square measure (Modigliani & Modigliani, 1997).

However, each of these measures has difficulties in addition to the proper computation of returns. For example, what exactly is the riskless rate of return for a given period? Analysts often use the six-month Treasury bill rate as a proxy, but what if the investor has a much longer investment horizon? The Jensen and Treynor measures require computation of market portfolio returns. What is the appropriate index for the market? Roll (1977) shows that different indices (even if highly correlated) will lead to very different performance rankings. Note that the Sharpe ratio does not require a market proxy. Once returns have been computed, exactly how are portfolio betas computed? This may be a significant problem, given that the composition of the portfolio is likely to constantly be in transition. Should portfolio beta be computed as a "time-weighted" average of individual security betas, or based on historical portfolio returns?

The following represent additional issues for investment institutions regarding the difficulties of using the above risk-adjusted portfolio performance measures:

1. Given that portfolio managers change jobs rather frequently, is it reasonable to measure fund performance rather than manager performance?
2. How frequently are we able to obtain enough data to obtain statistically significant measures of performance?
3. The CAPM, a model that decomposes return into compensation for time value of money and compensation for risk, serves as the basis for the Treynor and Jensen measures:

$$\text{CAPM: } E\left[r_p\right] = r_f + \left[\beta_p(r_m - r_f)\right]$$

   where $E[r_p]$, r_f, and r_m are the expected return on the portfolio, the riskless Treasury bill, and the market as a whole. However, the CAPM is only a single-time-period model. Multiple time periods and multiple cash flows cause problems in its application. In addition, many analysts will be concerned about the many assumptions that underlie the CAPM, as well as certain statistical tests that cast doubt on the empirical validity of the CAPM.
4. Investors holding funds representing only market segments might find that any measure based on the CAPM is inappropriate.
5. The Sharpe ratio will understate portfolio performance of undiversified portfolios in a setting where investors, in sum, hold numerous undiversified portfolios. That is, much of the risk captured in the Sharpe ratio can be diversified away. In addition, the Sharpe ratio is a better metric when portfolio returns are distributed normally. Skewness in returns will weaken the usefulness of the Sharpe ratio.
6. Errors in computing returns will bias measured betas downwards and will "slop" over into unsystematic variances (the part of risk that is unrelated to the market). Even seemingly minor problems can significantly bias beta measures. However, reasonably good correction procedures for betas measured with error do exist.

Portfolio Performance Benchmarking Illustration

Here, we will use the Jensen alpha measure to evaluate the performance of a portfolio:

$$J_p = \alpha_p = [r_p - r_f] - [\beta_p(r_m - r_f)]$$

Presumably, a positive alpha (α_p) that is statistically significant will indicate that the portfolio outperforms the market on a risk-adjusted basis. Consider Table 10.2, which records returns over a 20-year period for a hypothetical portfolio (p) and the market along with riskless return rates.

To compute the Jensen alpha measure, we first convert returns to risk premiums by subtracting riskless rates from returns on the portfolio and the market. We then run a regression of portfolio risk premiums against market risk premiums. The following represent the regression results and diagnostics:

$$[r_{p,t} - r_{f,t}] = \alpha_p + \beta_p(r_{m,t} - r_{f,t}) + e_t = \underset{(8.5056)}{0.073509} + \underset{(10.7452)}{0.951512}(r_{m,t} - r_{f,t}) + e_t \quad r\text{-squared} = 0.865127 \quad \text{d.f.} = 18$$

The regression t-statistics are reported in parentheses below the regression coefficients. First, we can conclude that the portfolio beta (0.951512) is statistically significant at the 1% level. The portfolio has slightly less systemic (market-related) risk than the market itself. The r-squared value suggests that slightly more than 86% of this fund's risk premiums are driven by market risk premiums. But most importantly, we can conclude that Jensen's alpha is positive and statistically significant at the 1% level. Thus, the fund outperformed the market on a risk-adjusted basis for the reported period. Although we have concluded statistical significance for alpha, we should question the relevance of data as old as 20 years to evaluation of current fund management. Figure 10.1 provides a pictorial display of our results.

TABLE 10.2 Portfolio p and Market Performance

Year	r_p	r_m	r_f	$r_p - r_f$	$r_m - r_f$	Year	r_p	r_m	r_f	$r_p - r_f$	$r_m - r_f$
1997	0.14	0.05	0.03	0.11	0.02	2007	0.11	0.07	0.05	0.06	0.02
1998	0.11	0.03	0.02	0.09	0.01	2008	0.22	0.17	0.06	0.16	0.11
1999	0.04	0.05	0.02	0.02	0.03	2009	0.22	0.16	0.07	0.15	0.09
2000	0.16	0.11	0.03	0.13	0.08	2010	−0.01	−0.05	0.06	−0.07	−0.11
2001	0.03	−0.12	0.02	0.01	−0.14	2011	0.04	−0.08	0.05	−0.01	−0.13
2002	0.14	0.09	0.03	0.11	0.06	2012	0.28	0.21	0.06	0.22	0.15
2003	0.26	0.13	0.04	0.22	0.09	2013	0.22	0.11	0.07	0.15	0.04
2004	0.26	0.18	0.05	0.21	0.13	2014	0.21	0.11	0.08	0.13	0.03
2005	0.13	0.04	0.05	0.08	−0.01	2015	−0.04	−0.11	0.07	−0.11	−0.18
2006	−0.08	−0.11	0.04	−0.12	−0.15	2016	0.18	0.12	0.05	0.13	0.07

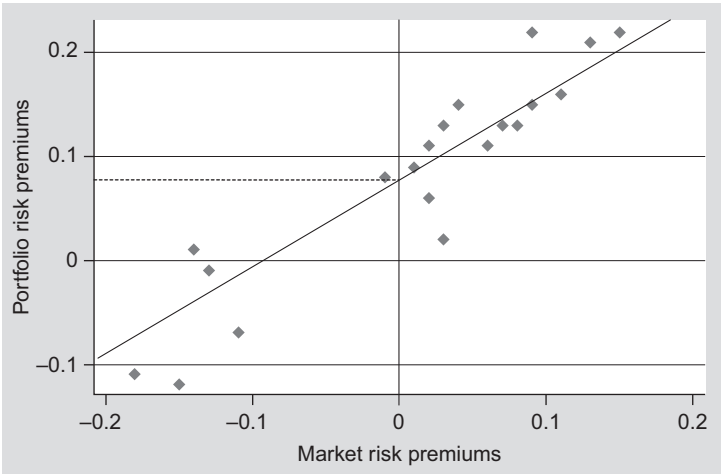


FIGURE 10.1 Jensen's alpha scatter diagram.

10.2 MARKET TIMING VERSUS SELECTION

Why might an investment portfolio earn consistently exceptional returns? First, higher risk portfolios tend to earn higher returns. This is why we benchmarked portfolios in the previous section. Second, strong portfolio performance might be due to simple luck. This is why we attempt to examine performance over many time periods and examine the statistical significance of our indicators. If an investment portfolio consistently earns higher than normal returns, it is useful to understand why. An important step in this process is to distinguish between skill at "security selection" and "market timing skill." Suppose, for example, an analyst is able to discern the future direction of the market. In periods before anticipated market upswings, the analyst should recommend increasing the beta of her portfolio to exploit this upswing. In the case of an anticipated market downturn, the analyst should recommend decreasing the beta of the portfolio.

The Quadratic Variable Approach

The CAPM concludes that there is a linear relationship between the risk premiums of well-diversified portfolios and risk premiums of the market portfolio, as in Figure 10.1. However, investors who correctly time the marker will shift their portfolios to lower betas during market downturns, shifting returns to be closer to the riskless rate. During market upturns, investors with strong timing ability will increase portfolio betas. Thus, for investors with strong market timing ability, the relationship between portfolio risk premiums and market risk premiums will be concave up as in Figure 10.2.

We return to our portfolio and market data from Table 10.2 in the previous section. We create an additional column for $(r_{m,t} - r_{f,t})^2$, which will serve as a second explanatory variable (see Treynor & Mazuy, 1966). This new testing variable accounts for the concave up relationship between portfolio and market risk premiums that strong timers will observe. We

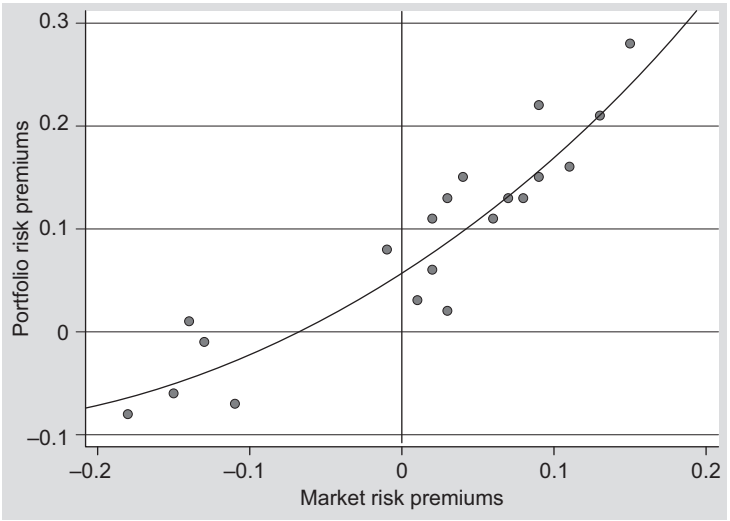


FIGURE 10.2 Jensen's alpha and timing ability.

now run a multiple regression to examine $[r_{p,t} - r_{f,t}]$ relative to $(r_{m,t} - r_{f,t})$ and $(r_{m,t} - r_{f,t})^2$. Our coefficient γ_p for $(r_{m,t} - r_{f,t})^2$ reflects timing ability on the part of the manager. A positive coefficient γ_p reflects the extent to which the manager shifts portfolios to reduce betas during periods of low market returns and the extent to which managers shift portfolio weights to increase beta during periods of high market returns. Since γ_p below is not statistically significant, this manager does not exhibit significant market timing ability. However, because Jensen's alpha is still positive and statistically significant, we conclude that this portfolio manager exhibits portfolio selection ability since overall performance seems superior.

$$\begin{aligned} r_{q,t} - r_{f,t} &= \alpha_p + \beta_p(r_{m,t} - r_{f,t}) + \gamma_p(r_{m,t} - r_{f,t})^2 + e_t \\ &= 0.0713 + 0.9605(r_{m,t} - r_{f,t}) + 0.2148(r_{m,f} - r_{f,t})^2 + e_t \\ &\quad (5.158) \quad (9.472) \qquad\qquad (0.201) \end{aligned}$$

Multiple adjusted r-squared = 0.85 d.f. = 17

Consider a second portfolio q with returns given in Table 10.3. Market returns and riskless rates are identical; however, we will see that returns for this second portfolio reflect beta shifting in particularly strong and weak markets. A similar quadratic regression is run for this second portfolio to check for this apparent timing ability, and results reveal significant timing ability. The plot for the second regression is pictured in Figure 10.2. Regression parameters are given by the following equation:

$$\begin{aligned} r_{q,t} - r_{f,t} &= \alpha_q + \beta_q(r_{m,t} - r_{f,t}) + \gamma_q(r_{m,t} - r_{f,t})^2 + e_t \\ &= 0.0580 + 1.0004(r_{m,t} - r_{f,t}) + 2.0425(r_{m,f} - r_{f,t})^2 + e_t \\ &\quad (4.350) \quad (10.230) \qquad\qquad (1.990) \end{aligned}$$

Multiple adjusted r-squared = 0.85 d.f. = 17

TABLE 10.3 Portfolio q and Market Performance

Year	r_q	r_m	r_f	$r_q - r_f$	$r_m - r_f$	Year	r_q	r_m	r_f	$r_q - r_f$	$r_m - r_f$
1997	0.14	0.05	0.03	0.11	0.02	2007	0.11	0.07	0.05	0.06	0.02
1998	0.05	0.03	0.02	0.03	0.01	2008	0.22	0.17	0.06	0.16	0.11
1999	0.04	0.01	0.02	0.02	0.03	2009	0.22	0.16	0.07	0.15	0.09
2000	0.16	0.11	0.03	0.13	0.08	2010	−0.01	−0.05	0.06	−0.07	−0.11
2001	0.03	−0.12	0.02	0.01	−0.14	2011	0.04	−0.08	0.05	−0.01	−0.13
2002	0.14	0.09	0.03	0.11	0.06	2012	0.34	0.21	0.06	0.28	0.15
2003	0.26	0.13	0.04	0.22	0.09	2013	0.22	0.11	0.07	0.15	0.04
2004	0.26	0.18	0.05	0.21	0.13	2014	0.21	0.11	0.08	0.13	0.03
2005	0.13	0.04	0.05	0.08	−0.01	2015	−0.01	−0.11	0.07	−0.08	−0.18
2006	−0.02	−0.11	0.04	−0.06	−0.15	2016	0.18	0.12	0.05	0.13	0.07

Coefficients α_q and β_q are statistically significant at the 1% level and coefficient γ_q is significant at the 6% level. Thus, portfolio q reflects both superior overall performance ($\alpha_q = 0.0580$) and likely superior timing ability ($\gamma_q = 2.0425 > 0$). The upward concavity of the relationship between $(r_{q,t} - r_{f,t})$ and $(r_{m,t} - r_{f,t})$ suggests that the investor tends to shift her portfolio beta down (and portfolio returns up towards the riskless return rate) in weak markets, and that she tends to shift her portfolio beta up (and portfolio returns up further away from the riskless return rate) in strong markets. However, in a quadratic regression such as this, the analyst needs to be careful about interpreting α_q. For example, a negative or statistically insignificant α_q in this quadratic regression should not be interpreted to imply investment performance inferior to or on par with the market, particularly if γ_q is positive and statistically significant.

The Dummy Variable Approach

An alternative indicator of market timing is based on a regression including an interaction term, which is a combination of a dummy variable D (takes on a value of 0 or 1) and market risk premiums:

$$r_{q,t} - r_{f,t} = \alpha_q + \beta_q(r_{m,t} - r_{f,t}) + \gamma_q D(r_{m,t} - r_{f,t}) + e_t$$

$$\text{where: If } (r_{mt} - r_{ft}) \geq 0,\ D = 0$$

$$\text{If } (r_{mt} - r_{ft}) < 0,\ D = 1$$

Thus, if the market return equals or exceeds the riskless return, β_q is the portfolio beta; that is, β_q can be taken to be the up-market beta. If the market return is less than the riskless return, $\beta_q - \gamma_q$ can be taken to be the portfolio beta; that is, $\beta_q - \gamma_q$ is the down-market beta. The portfolio's shift in beta is γ_q, so that a negative and statistically significant γ_q is

indicative of market timing ability. For our numerical illustration for portfolio q, multiple regression results are reported as follows:

$$\begin{aligned} r_{q,t} - r_{f,t} &= \alpha_q + \beta_q(r_{m,t} - r_{f,t}) + \gamma_q D(r_{m,t} - r_{f,t}) + e_t \\ &= 0.050 + 1.311(r_{m,t} - r_{f,t}) - 0.672D(r_{m,t} - r_{f,t}) + e_t \\ &\quad (2.931)\ (5.853) \qquad\qquad (-1.928) \end{aligned}$$

Coefficients α_q and β_q once again are statistically significant at the 1% level and coefficient γ_q is significant at the 6% level. Thus, portfolio q reflects superior overall performance ($\alpha_q = 0.050$) and likely superior timing ability ($\gamma_q = -0.672 < 0$). It appears that the portfolio manager shifts the beta of her portfolio down by 0.672 in down markets relative to up markets. Again in this regression with the interactive term, the analyst needs to be careful about interpreting α_q. For example, a negative or statistically insignificant α_q in this type of regression should not be interpreted to imply investment performance inferior to or on par with the market, particularly if γ_q is positive and statistically significant.

10.3 TRADE EVALUATION AND VOLUME-WEIGHTED AVERAGE PRICE

The tools that we discussed in the previous two sections are particularly appropriate for evaluating portfolios and portfolio strategies over longer periods of time. Their focus is on overall portfolio performance, without distinguishing the trading aspect of the investment decision from the selection and timing aspects. Evaluating trader performance independently of portfolio performance is important for trading firms, portfolio managers, funds, and other financial institutions. Unfortunately, evaluating trader performance is not a simple task. The trading decision is different from the investing or portfolio management decision, and needs to be evaluated accordingly. We discussed portfolio benchmarking earlier; here, we discuss trade benchmarking. Trade benchmarking is generally based on comparing a trade or series of trades against the hypothetical results of a series of trades based on a specific or benchmark strategy.

Traders are expected to seek the "best execution" for their transactions. But what is best execution? As we discussed earlier in the book, this concept is rather vague. The Chartered Financial Analyst (CFA) Institute's trade-management guidelines (CFA, 2004) characterize best execution "as the trading process firms apply that seeks to maximize the value of a client's portfolio within the client's stated investment objectives and constraints." Obviously, this characterization is ambiguous and difficult to quantify and its complexity leads to many problems. More generally, deciding on an appropriate evaluation metric or benchmarking strategy is a difficult decision. Many firms use one or more of the following:

- *Volume-Weighted Average Price* (VWAP, introduced by Berkowitz, Logue, & Noser, 1988): Calculated by dividing the dollar volume of trading in a stock by the share volume over a given period of time, typically one day.

- *Arrival Price*: The midpoint of the bid−offer spread at the time the order is received (bid-ask midpoint or BAM). See also Section 6.4, Adverse Selection and the Spread.
- *Market-on-Close* (MOC): The last price obtained by a trader at the end of the day relative to the last price reported by the exchange.
- *Implementation shortfall*: The performance difference between the hypothetical profits realized by a paper or theoretical portfolio replicating an actual portfolio ignoring friction costs and the profits realized by the actual portfolio.

VWAP

VWAP, introduced by Berkowitz et al. (1988) and popularized by the trading consulting firm Abel/Noser, is the most commonly used of the trading evaluation metrics and benchmarks. VWAP is the average price at which a stock trades in a given period, weighted by the volume traded at each price. Suppose that a customer engaged a broker to purchase a large number of shares. Since the customer has already made the decision to invest in the shares, the broker should be evaluated only on the trade execution. VWAP can be used as a benchmark to evaluate the quality of the execution provided by the broker. If the brokerage firm's purchases were made at a lower VWAP than the market VWAP for the relevant period, we might infer that the firm handled the order well for the customer. VWAP, either for the trader or for the market, is calculated as follows:

$$VWAP = \frac{\sum_j Q_j P_j}{\sum_j Q_j} \tag{10.7}$$

where Q_j represents the number of shares executed in a given order j at a price of P_j. Thus, VWAP is the transaction execution average price of over a given period of time. While VWAP is usually computed for the day of the trade, sometimes traders use multiday VWAP. This is likely when orders broken up for execution over several days. In addition, intraday VWAPs can be used for orders executed entirely within a given trading day or some portion thereof. A trader (broker) might be regarded to have abnormal trading performance if she can consistently beat the market or other relevant VWAP on her trades.

One criticism of VWAP cost is that trader's own trades will be included in the benchmark price. This is corrected by excluding from VWAP the trader's own transactions, although it can cause problems when the trader's own trades account for a significant portion of the total. Suppose, for example, that the trader was counterparty to all security trades for that period. Obviously, VWAP would not be useful. But, this issue can arise whenever any illiquid security is being traded. One problem with performance evaluation based on daily VWAP benchmarks is that this benchmark can encourage traders to spread their trades out over the day simply avoid the risk of trading at the high or low prices for the day.

VWAP is quick and easy to compute, understand, and compare. This simplicity is VWAP's main strength. In addition, VWAP is also useful for securities that trade in different markets. However, VWAP becomes less meaningful when the security's price trends up or down over the course of the trading period. In addition, VWAP does not account for costs related to trade delays, trade cancellations or failures to execute trades. In fact,

VWAP is not useful when completing the transaction immediately is urgent. An immediate execution requirement does not allow the trader any discretion on which to be evaluated. Nor does VWAP account for slippage, the trader's price impact on the security. Executing a large transaction at once can certainly impact the security price, but this impact will not necessarily adversely affect the broker's apparent performance against VWAP. In fact, a trader might even be able to manipulate this price impact and VWAP by increasing the volume of his trades along with the impact on VWAP.

Traders can practically always match VWAP by merely executing their transactions at the same prices and with the same proportional shifts in volume as the market over the transaction period. Forecasting volume is important for traders benchmarking against VWAP. For example, in U.S. markets, stock trade volume tends to be higher earlier in the morning and later in the afternoon sessions with lower volumes around midday. In European markets, volumes tend increase in afternoon hours when U.S. markets are open.

VWAP: A Simple Illustration

Suppose that a broker has been instructed to purchase 600 shares at the market. She does so, purchasing them for a total price of 30,011. Market transactions over the security's intraday execution period are listed in Table 10.4. The broker's executed transactions were the second through fourth transactions in the table. The total volume of shares exchanged was 2500, with a total value of 125,098. Hence, VWAP for these transactions was 125,098/2500 = 50.0392. The broker purchased 100 shares in the second transaction at $50.01, 400 shares in the third transaction at $50.02, and 100 shares in the fourth transaction at $50.02. The average share price paid by the broker was 30,011/600 = 50.0183. Our calculations suggest that the broker beat the market VWAP. However, note that she did so by purchasing early or quickly in a rising market.

TABLE 10.4 Price and Volume for VWAP Calculations

Price ($)	Volume	$P \times Q$
50.01	200	10,002
50.01	100	5001
50.02	400	20,008
50.02	100	5002
50.03	200	10,006
50.05	600	30,030
50.06	400	20,024
50.05	500	25,025
Totals	2500	125,098

10.4 IMPLEMENTATION SHORTFALL

A delay will likely occur between the time an investor, portfolio manager or broker decides to execute a transaction and when the transaction is actually executed. This delay may increase the cost of the transaction to the investor or portfolio manager. *Implementation shortfall* can be defined as the difference in prices between the ideal transaction and the actual implemented transaction. The ideal transaction price is typically based on the security price existing when the transaction decision was made, or in the case of an agency broker, at the order arrival. The difference between this decision price and the actual implemented or execution price is construed to be implementation shortfall. Implementation shortfall is indicative of the quality of the trade execution.

Suppose that a trader or portfolio manager makes a decision to buy or sell a security based on its current price, referred to here as its decision price. Similarly, an agency broker might instead use for her own trade execution evaluation the arrival price (e.g., bid/ask midpoint), the price of the security when she actually receives the order. However, since the execution price of an order normally differs from the original decision or arrival price, the actual implementation of an investment strategy by the trader, broker or portfolio manager leads to four primary types of friction or implementation costs. Implementation shortfall can be decomposed into these four types of friction costs:

1. Broker, exchange, and other explicit fees and commissions. Frequently, brokers bundle exchange, SEC, and other fees into their own commissions. Small transactions tend to have higher proportional explicit transactions costs.
2. Delay costs, based on the price difference between the portfolio manager's decision price and the broker's arrival price.
3. Price impact costs associated with transaction executions (slippage). Buy orders will exert upward price pressure on the security; sell orders will exert downward pressure. Larger transactions will tend to have larger impact costs.
4. Opportunity costs associated with transactions; that is, the opportunities and profits that were forgone prior to the trade's completed execution. Opportunity costs can also include the portion of an order that was canceled due to an unfulfilled limit order restriction.

We will seek to analyze and segregate these four types of costs, which will sum to implementation costs. First, to an extent, the costs of market impact, which increase as the execution speed of the transaction increases, balance against delay and opportunity costs, which increase as complete execution of the transaction is delayed. When paper portfolios are created for back-testing and other purposes, prices for transactions are taken from actual bid and offer quotations or from actual transactions taken from historical market data. These paper transactions do not normally include the costs of implementing the portfolio strategies. Implementation shortfall is the performance difference between the hypothetical profits realized by a paper or theoretical portfolio replicating an actual portfolio ignoring friction costs and the profits realized by the actual portfolio. That is, implementation shortfall is the difference between the decision or order arrival price and the realized price of the execution, including broker and exchange fees, slippage, and transaction opportunity costs.

It can be difficult to distinguish these four types of implementation costs. Individual judgment is likely to play a role in segregating these costs. Brokerage, market and other explicit costs are usually straightforward. Opportunity costs can also be straightforward if their definition is limited to portions of orders canceled due to non-execution, perhaps due to order limits.

Illustration: Implementation Shortfall

Suppose that a portfolio manager makes a decision to purchase 10,000 shares of a fairly illiquid stock one hour before its open, based on its $50.00 closing price the prior day (the decision or ideal price), and places a limit order to buy for 50.45. The stock opened at 9:30 at a price of 50.20, and 1000 shares are purchased at 9:31 at a price of 50.25. 50.25 can be considered to be the broker's arrival price if orders arriving earlier had priority at the open. At 9:32, 5000 shares are purchased for 50.40, and 1000 more for 50.30 at 11:03. An additional 1000 shares are purchased for 50.30 at 2:15, the market price quickly rises to 50.48 and closes at 50.50 with 2000 shares in the order unexecuted. These transactions are represented in Table 10.5. The commissions, including all explicit fees, were $0.01 for each of 8000 shares. The total implementation cost is $3980.

The ideal portfolio profit, based on 10,000 shares, the $50 decision price and the following day closing price, is $5000 = $505,000−$500,000. However, actual executions, summarized in Table 10.5, reveal an actual portfolio cost of $402,980 = $50,250 + $252,000 + $50,300 + $50,350 + $80. The end-of-following day portfolio value based on 8000 shares is $404,000, yielding an actual profit of $1020 = $404,000−$402,980 and an implementation shortfall of $3980 = $5000−$1020. Table 10.6 breaks down the implementation shortfall into its component costs.

Now, we will decompose the implementation shortfall into its various components. First, the delay to the first execution cost 0.25 per share, affecting all 10,000 shares for a total of $2500 in delay costs. The 11:03 transaction had additional per share delay costs of 0.05 for each of 1000 shares relative to the 9:31 transaction, and the 2:15 transaction had additional delay costs of $0.10 for each of 1000 shares, leading to a total of

TABLE 10.5 Implementation Shortfall Executions

t	Volume	Price$_t$
Close		50.00
Open		50.20
9:31	1000	50.25
9:32	5000	50.40
11:03	1000	50.30
2:15	1000	50.35
Close		50.50

TABLE 10.6 Implementation Shortfall

Commissions	80
Delay cost	2650
Price Impact	750
Opportunity costs	500
Total costs	3980

$\$2500 + 50 + 100 = \$2{,}650$ in total delay costs. The 5000 share transaction executed at 9:32 was particularly large and seemed to exert upward price pressure, as the price rose from 50.25 to 50.40 in about a minute. Potential price pressure is ignored on other transactions because it did not seem pronounced, and the transaction execution were sufficiently delayed relative to previous transactions to not be related. The price impact cost of this 9:32 transaction is estimated at $5000 \cdot (50.40 - 50.25) = \$750$. Finally, 2000 shares failed to execute, with a "virtual" opportunity cost of $2000 \times (50.50 - 50.00 - 0.25) = \$500$, where 0.25 is the associated delay cost already accounted for. Thus, Table 10.6 sums the implementation shortfall of the executions. This breakdown of the explicit and implicit implementation costs illustrates the typical magnitude of the implicit costs relative to the explicit transaction costs. To some extent, it can be difficult to assign costs to the various components, and different analysts, depending on the circumstances, may assign costs differently.

10.5 VALUE AT RISK

Value at Risk (*VaR*) is a standard tool for risk management in financial institutions. It provides a figure for the loss in a portfolio over a given time frame, with a given set of distributional assumptions at a given confidence level. In other words, in typical usage, if asset values are distributed normally with a given variance, the analyst can calculate a loss figure that will not be exceeded at a specified level of confidence. Thus, for example, with given normally distributed asset values, the analyst can calculate a "worst case scenario" at a given confidence level, typically 95% or 99%. For a confidence level of $\alpha = 99\%$, one can be 99% certain that at the end of the given time horizon there will be no greater loss than just the *VaR*—as long as the model's assumptions hold true. Thus, the *VaR model* concerns the probability that a loss of a given size will not be exceeded over a given period of time:

$$\begin{aligned} VaR = {} & Asset\ Value \times Daily\ Return\ Standard\ Deviation \\ & \times Confidence\ Interval\ Factor \times the\ Square\ Root\ of\ Time \\ VaR = {} & Asset\ Value \times \sigma \times z \times \sqrt{t} \end{aligned} \tag{10.8}$$

Asset value is the total value of the institution, or its relevant unit, which might even be the portfolio of a single trader or desk. The daily (or some other time interval) return

standard deviation applies to this asset value, the confidence interval factor represents the maximum acceptable probability that this loss will be exceeded (typically a z-value such as 1% from a normal distribution), and time is measured in days or some other period.

Derivatives Portfolio *VaR*: Illustration

Suppose a bank with $1 billion in its derivatives portfolio experiences a 0.5% standard deviation in its normally distributed daily returns. The one-tailed z-value corresponding with the 1% confidence interval is 2.326. The bank seeks to determine the size of a loss that has a 1% probability of being incurred over a five-day week:

$$VaR = \$1,000,000,000 \times 0.005 \times 2.326 \times \sqrt{5} = \$26,005,471$$

Thus, assuming that daily asset returns are normally distributed (this is an extremely important assumption) with a standard deviation of 0.005, there is a 1% probability that the bank will experience a loss exceeding $26,005,471. While this figure is fairly small compared to the total asset level, its importance is enhanced by the bank's borrowing levels. For example, if the bank were 30-to-1 debt financed, its initial equity level would be $33,333,333, and this loss would wipe out almost 80% of the bank's equity stake. This $26,005,471 *VaR* figure would normally be regarded to be extremely high given the level of this bank's equity capital. Thus, if bank profits are normally distributed, 99% of the time, losses realized by the institution will be less than the computed 5-day *VAR* figure. However, if the bank's profits were not normally distributed as assumed, but instead were subject to a distribution with more extreme observations than implied by normal distributions (fat tails), equity capital might have a significantly higher probability of being wiped out after five days.

Calculating, using, and interpreting *VaR* for a single asset with a known volatility and normal distribution of returns is fairly straightforward. As we will demonstrate shortly, the same is true for a portfolio of similarly "well-behaved" assets. However, portfolios of derivative securities or significantly different assets can be more of a problem. For example, the returns of put and call options are not normally distributed. Calculating *VaR* for a larger mix of options, even on a single security, can be quite difficult. Some traders will focus on deltas and gammas of portfolio components, but even these are problematic when underlying security variances are large. Recall from Chapter 9 that the "Greeks" only apply when underlying security price changes are small. But changes required to be consistent with 95% or 99% confidence intervals are normally large. In such instances, *VaR* calculations can be facilitated with the use of Monte-Carlo simulations. Again, as we alluded to above, working with appropriate and realistic distributions is essential to interpreting *VaR*.

Issues Arising with *VaR* Applications

VaR assumptions, methodology, and applications have been heavily criticized, in large part because of *VaR*'s reliance on potentially unrealistic assumptions and giving traders

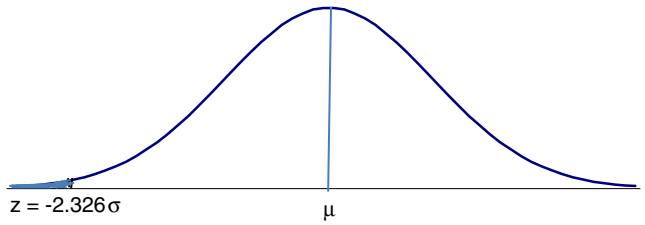


FIGURE 10.3 Bottom 1% area under the normal curve.

and institutions false hopes concerning their ability to measure and control risk. Perhaps the most important of these criticisms concerns the assumption of normally distributed returns. Notice in Figure 10.3 the depiction of the narrow tails in the normal distribution, suggesting that the most disastrous outcomes (sometimes referred to in a risk measurement context as black swans) are very unlikely (see Taleb [2010]). If worst-case or near-worst-case scenarios are not actually rare, those extreme situations might instead be parts of *"fat tails,"* and their probabilities might be underestimated by the thin tails consistent with a normal distribution. Thus, do consider the implications of the normal distribution before relying too heavily on the assumption of normality. Remodel with another distribution if necessary. Similarly, if standard deviation is not the best or most complete measure of variability of returns, or if variability of returns does not provide a good indicator of risk, then adapt the *VaR* methodology to a more appropriate measure of risk, and/or rely on other risk measurement methods.

VaR and *SPAN*

As discussed in Chapter 3, the Chicago Mercantile Exchange (CME) developed and implemented its *Standard Portfolio Analysis of Risk* (*SPAN*) model in 1988 to evaluate overall portfolio risk for its clients and other stakeholders. SPAN, and its updated generations such as SPAN2, are market simulations based on *VaR* methods and have served as widely used performance bond mechanisms or risk-based margin requirement calculations for a wide variety of combinations of securities and instruments. Older SPAN simulations typically calculated security and portfolio outcomes based on 16 potential "what-if" risk scenarios or hypothetical shocks in a grid. Newer simulations account for security correlations and many more scenarios. The outcomes of the simulations, conducted at least once per day by CME, can be used to evaluate the risk positions of relevant securities, combinations of securities, portfolios, and institutions themselves and to calculate performance bond (margin) requirements for CME clients, including exchanges and clearinghouses throughout the world.[2] Other exchanges have developed or are developing somewhat similar products, including Prisma (Eurex) and PAIRS (LCH).

[2]For details, see CME Group (2019).

Equity Portfolio *VaR*: Illustration

Here, we illustrate a simple *VaR* calculation for an equity portfolio made up of three stocks. For sake of simplicity, we will assume that portfolio returns are normally distributed. The trader has "borrowed" $900,000 from his employer and invested $100,000 of his "own" money, for a total of $1,000,000. This trader's employer requires that the trader's one-week portfolio *VaR* not exceed his trading capital of $100,000, with a 99% degree of confidence. We will examine the trader's portfolio and its three component stocks to determine whether he is within his risk limits.

Table 10.7 lists the three securities, 1, 2, and 3, along with their portfolio weights, annualized return variances, and covariances with each other. We calculate the portfolio variance with either Equations (10.9), (10.10), or (10.11). Equation 10.9 is the general expression for an n-security portfolio variance. Equation 10.10, written for $n = 3$ securities, combines terms from Equation 10.9. Equation 10.11 is simply Equation 10.9, written in matrix format for $n = 3$ securities. Equation 10.11 might be the easiest to set up on a spreadsheet using the matrix multiplication calculation functions (see the Mathematics Appendix A.2 near the end of the text for matrix computation details). We find that portfolio variance equals 0.0904, and its standard deviation of returns is 0.3. Weekly *VaR* is calculated with Equation (10.8) within a 99% confidence interval, using a "z" value of 2.326, and $t = \sqrt{1/52}$ years. This *VaR* calculation suggests that the trader's portfolio does lie within his 99% confidence interval *VaR* risk limits for one week.

$$\sigma_p^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} w_i\, w_j\, \sigma_{ij} \tag{10.9}$$

$$\begin{aligned}\sigma_p^2 &= (w_1^2 \cdot \sigma_1^2) + (w_2^2 \cdot \sigma_2^2) + (w_3^2 \cdot \sigma_3^2) + 2(w_1 \cdot w_2 \cdot \sigma_{1,2}) + 2(w_1 \cdot w_3 \cdot \sigma_{1,3}) + 2(w_2 \cdot w_3 \cdot \sigma_{2,3}) \\ &= (0.04 \cdot 0.04) + (0.25 \cdot 0.16)(0.09 \cdot 0.36) + 2(0.2 \cdot 0.5 \cdot 0.01) + 2(0.2 \cdot 0.3 \cdot 0.02) \\ &\quad + 2(0.5 \cdot 0.3 \cdot 0.04) = 0.0904\end{aligned} \tag{10.10}$$

$$\sigma_p^2 = \begin{bmatrix} w_1 & w_2 & w_3 \end{bmatrix} \begin{bmatrix} \sigma_1^2 & \sigma_{1,2} & \sigma_{1,3} \\ \sigma_{2,1} & \sigma_2^2 & \sigma_{2,3} \\ \sigma_{3,1} & \sigma_{3,2} & \sigma_3^2 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} 0.2 & 0.5 & 0.3 \end{bmatrix} \begin{bmatrix} 0.04 & 0.01 & 0.02 \\ 0.01 & 0.16 & 0.04 \\ 0.02 & 0.04 & 0.36 \end{bmatrix} \begin{bmatrix} 0.2 \\ 0.5 \\ 0.3 \end{bmatrix} = 0.0904 \tag{10.11}$$

$$VaR = Asset\ Value \times \sigma \times z\sqrt{t} = \$1,000,000 \times \sqrt{0.0904} \times 2.326 \cdot \sqrt{1/52} = \$96,996.84 \tag{10.12}$$

TABLE 10.7 Portfolio Weights and Security Characteristics

i	w_i	σ_i	$\sigma_{1,i}$	$\sigma_{2,i}$	$\sigma_{3,i}$
1	0.20	0.20	0.04	0.01	0.02
2	0.50	0.40	0.01	0.16	0.04
3	0.30	0.60	0.02	0.04	0.36

We infer from this *VaR* calculation that there is a 99% chance that the loss on this million dollar portfolio will not exceed $96,996.84, or approximately 97% of the trader's equity capital.

Additional Reading

The CFA Institute's (2004) *Trade Management Guidelines* is helpful for better understanding of the ambiguous concept of "best execution," and the need for and how to establish trade execution policies, disclosures, and record-keeping. In addition, issues related to trade performance have appeared on recent CFA Level III exams and will appear on blogs and study sites related to the CFA Level III. The portfolio management textbook by Elton et al. (2010) provides an excellent introduction to portfolio evaluation techniques in chapter 25. Perold (1988) provides a readable practitioner-oriented introduction to implementation shortfall, and Hasbrouck (2007) discusses implementation shortfall and trading costs in chapters 14 and 15. Jorion (1997) offers a standard and comprehensive introduction to *VaR*. Harris (2003) offers excellent reading on transaction costs and performance evaluation in chapters 21 and 22. Wagner (2008) wrote a highly readable account of his personal experiences as a consultant in performance evaluation and with transaction costs. Chapter 9 of Gregory (2014) provides a nice and somewhat more technical description of various margin and default fund methods, including *VaR* and its applications, particularly in a derivatives context.

References

Berkowitz, S. A., Logue, D. E., & Noser, E. A. (1988). The total costs of transacting on the NYSE. *Journal of Finance, 43*, 97–112.

CFA Institute. (2004). *Trade management guidelines*. Available at: <http://www.cfapubs.org/doi/pdf/10.2469/ccb.v2004.n3.4007> Accessed 25.03.11.

CME Group. (2019). CME SPAN®: Standard Portfolio Analysis of Risk®. Unpublished slides. Available at: <http://www.cmegroup.com/clearing/files/span-methodology.pdf> Accessed 10.20.21.

Elton, E., Gruber, M., Brown, S., & Goetzman, W. (2010). *Modern portfolio theory and investment analysis* (8th ed.). New York: John Wiley.

Gregory, J. (2014). *Central counterparties: Mandatory clearing and bilateral margin requirements for OTC derivatives*. Chichester, UK: Wiley.

Harris, L. (2003). *Trading and exchanges: Market microstructure for practitioners*. Oxford: Oxford University Press.

Hasbrouck, J. (2007). *Empirical market microstructure: The institutions, economics and econometrics of securities trading*. New York: Oxford University Press.

Jensen, M. (1968). The performance of mutual funds in the period 1945–1964. *Journal of Finance, 23*(2), 389–416.

Jorion, P. (1997). *Value at risk*. Chicago: Irwin Press.

Modigliani, F., & Modigliani, L. (1997). Risk-adjusted performance. *Journal of Portfolio Management, Winter*, 45–54.

Perold, A. (1988). The implementation shortfall: Paper versus reality. *Journal of Portfolio Management, 14*(3), 4–9.

Roll, R. (1977). A critique of the asset pricing theory's tests, Part I: On past and potential testability of the theory. *Journal of Financial Economics, 4*(2), 129–176.

Sortino, F. (1996). On the use and misuse of downside risk. *Journal of Portfolio Management, 22*(2), 35–42.

Taleb, N. N. (2010). *The Black Swan: The impact of the highly improbable* (2nd ed.). London: Penguin.

Treynor, J. L., & Mazuy, M. (1966). Can mutual funds outguess the market? *Harvard Business Review, 44*(4), 131–136.

Wagner, W. H. (2008). The incredible story of transaction cost management: A personal recollection. *Journal of Trading, 3*(3), 8–14. Available at: <http://www.scribd.com/doc/6419002/Incredible-Story-of-Transaction-Cost-Measurement> Accessed 25.03.11.

10.6 EXERCISES

1. On February 1, a hedge fund initiated activities with a trading account of $1,000,000 and 20,000 outstanding shares of stock. On January 2, the fund realized $30,000 in trading profits. On February 3, the fund again realized $30,000 in trading profits, but paid a dividend of $30,000. On February 4, the fund incurred $50,000 in trading losses. Trading profits for February 5 were $10,000. In addition, on February 5, investors redeemed 1000 shares at the NAV.
   a. What were NAVs for the fund for each day of the week beginning February 1?
   b. What were daily returns for February 2 through February 5?
   c. What was the time-weighted average return for this four-day period?
2. Suppose that we wished to compute monthly returns for Fund A over a period of five months. We collect relevant end-of-month NAVs along with any dividends. The following table lists NAV's and dividends collected for Fund A from June 30 to November 30. Following the NAV data are sample return calculations:

Date		NAV_t	NAV_{t-1}	DIV_t
June 30	1	50	—	0
July 31	2	55	50	0
Aug. 31	3	50	55	0
Sep. 30	4	54	50	0
Oct. 31	5	47	54	2
Nov. 30	6	51	47	0

   a. Calculate returns for each month starting with July.
   b. The following lists NAVs for Fund B. Calculate monthly returns for Fund B.

Fund B

Date	t	NAV_t	NAV_{t-1}	DIV_t
June 30	1	50	—	0
July 31	2	80	50	0
Aug. 31	3	40	80	0
Sep. 30	4	60	40	0
Oct. 31	5	30	60	0
Nov. 30	6	45	30	0

   c. Calculate geometric mean returns for Funds A and B.
   d. Calculate time-weighted average returns for Funds A and B.
3. The following table provides historical *percentage* returns for the Patterson and Liston Funds along with percentage returns on the market portfolio (index or fund):

Year	Patterson	Liston	Market
2012	4	19	15
2013	7	4	10
2014	11	−4	3
2015	4	21	12
2016	5	13	9

Suppose that the riskless rate of return (or T-Bill rate) was 3% for each year. Calculate the following based on the preceding table:

a. Mean historical returns for the two funds and the market portfolio.
b. Variances associated with Patterson Fund returns and Liston Fund returns along with returns on the market portfolio.
c. The historical covariance and coefficient of correlation between returns of the Patterson Fund and returns on the market portfolio.
d. The historical covariance and coefficient of correlation between returns of the Liston Fund and returns on the market portfolio.
e. Historical betas for Patterson and Liston Funds, based on five years of returns data. (Note that the data sets are too limited in size to be considered very reliable.)
f. Suppose that a market return of 8% was obtained in the year 2017, and that the riskless return for 2017 was 3%. Further suppose that the Patterson and Liston Funds earned 10% and 14%, respectively, over 2017. Based on a Treynor Index, how did each of these funds perform relative to the market?
g. Comment on the 2017 risk-adjusted performance for each of the two funds based on Treynor Index results.

4. Historical returns for the Ripco Fund and the market portfolio along with Treasury bill (T-Bill) rates (r_f) are summarized in the following table:

Year	Ripco	Market	T-Bill	Year	Ripco	Market	T-Bill
1	0.35	0.28	0.05	11	−0.07	−0.15	0.05
2	0.04	0.05	0.05	12	−0.11	−0.21	0.05
3	0.10	0.09	0.05	13	0.42	0.26	0.05
4	−0.01	−0.02	0.05	14	0.01	0.04	0.05
5	0.38	0.32	0.05	15	0.05	0.08	0.05
6	0.31	0.25	0.05	16	0.07	0.11	0.05
7	0.33	0.26	0.05	17	−0.01	−0.03	0.05
8	0.42	0.30	0.05	18	−0.12	−0.37	0.05
9	0.07	0.14	0.05	19	0.33	0.30	0.05
10	−0.02	−0.06	0.05	20	0.05	0.13	0.05

a. Calculate the fund beta over the 20-year period.
b. Calculate Jensen's alpha for the fund over the 20-year period. Did the fund outperform the market during this period based on Jensen's alpha?
c. Based on a squared market risk premium term, does the Ripco Fund demonstrate market timing ability?
d. Based on an appropriate interaction term, does the Ripco Fund demonstrate market timing ability?

5. Many analysts argue that the Sharpe ratio is a more appropriate performance metric for an investor's overall portfolio than for a given segment or sector of her portfolio. Prepare a statement to support this perspective.

6. Suppose that the market for Stock X on a given date consisted of the transactions listed in the following table. The left two columns represented transactions for Stock X, excluding the broker's orders. The right two columns consisted of the broker's orders only, all of which were orders to buy. What was the VWAP for shares of X stock for that day? What was the VWAP excluding the broker's own orders? How did the broker perform relative to the market for his buy transactions?

Market for X Excluding Broker's Transactions		**Broker's Transactions for X**	
# Shares	**Price**	**# Shares**	**Price**
30,000	100.00	10,000	100.00
40,000	100.02	10,000	100.01
10,000	100.05	10,000	100.03
20,000	100.06	20,000	100.04
30,000	100.07	40,000	100.05
10,000	100.09	40,000	100.05

7. **a.** On what factors does implementation shortfall depend? That is, what affects the various types of implementation costs?
b. Why might increasing the speed at which a large transaction is executed increase the implementation shortfall of an order?
c. Why might decreasing the speed at which a large transaction is executed increase the implementation shortfall of an order?

8. A fund manager decided to purchase 1000 shares of General Statics stock after the market closed with a price of 20. The fund manager placed a limit order at 20.15 to buy 1000 shares before the market opened on the following day. The market opened at 20.10. 400 shares of the limit order quickly executed at 20.10 and 400 more at the limit order price of 20.15. The remaining 200 shares of the order failed to execute and the market closed at 20.20. The commission paid to the broker was 0.04 per share.

a. Calculate the implementation shortfall for this order.
b. Decompose implementation shortfall into its commissions, delay costs, price impact costs, and opportunity costs components.
c. Is it possible for implementation costs to be negative? Why or why not?

9. A small trading firm has a securities portfolio with a current value equal to $80,000,000. The portfolio's returns tend to be normally distributed with a 1% daily standard deviation and uncorrelated over time. The firm will use its *VaR* calculation to produce the maximum value of the loss that it is willing to incur on its portfolio at several thresholds. In particular, the firm seeks to calculate its portfolio *VaR* at 95% and 99% over a single day and over a 30-day month. That is, the firm seeks to calculate the maximum loss that it can incur at these probability thresholds. What are these four threshold values? How might these *VaR* figures be interpreted?

10. The *VaR* method can be applied to individual or institutional investment portfolios. For example, suppose that you have a $100,000 asset portfolio for which you seek to compute *VaR*. The portfolio experiences a 1% weekly standard deviation on its returns. What is the lower-end range size of a loss that occurs with 5% probability over a 4-week month?

11. What are the most essential criticisms of the *VaR* method as an indicator of stress of a financial institution (or unit thereof)?

12. An investor has combined securities X, Y, and Z into a portfolio. He has invested $1000 in Security X, $2000 into Security Y, and $3000 into Security Z. Security X has an expected return of 10%; Security Y has an expected return of 15%; and security Z has an expected return of 20%. The standard deviations associated with Securities X, Y, and Z are 12%, 18%, and 24%, respectively. The coefficient of correlation between returns on Securities X and Y is 0.8; the correlation coefficient between X and Z returns is 0.7; and the correlation coefficient between Y and Z returns is 0.6.
a. Find the expected return and standard deviation of the resultant portfolio.
b. Over the course of one week, what is the *VaR* of this 3-security portfolio at the 1% confidence level?

CHAPTER

11

The Mind of the Investor

11.1 RATIONAL INVESTOR PARADIGMS

As we will discuss later in Chapter 12, institutional buy-side investors such as mutual funds generally fail to outperform the market on a risk-adjusted basis. Many individual investors such as day traders tend to perform very poorly as well, perhaps even worse than their institutional buy-side counterparts. Barber et al. (2020) note that less than 3% of active day traders can produce predictably positive net returns. Nevertheless, the authors point out that the most experienced unprofitable day traders continue to trade and continue to lose. In fact, their study suggests that nearly three-fourths of day trading can be traced to traders with histories of trading losses.

So, are most day traders just plain stupid? Grinblatt, Keloharju, and Linnainma (2011), based on a sample of nearly 160,000 Finnish Armed Forces intelligence tests taken between 1982 and 2000 along with tax returns data, found that stock market participants had higher IQs than their noninvesting counterparts, even after adjusting for income and wealth differences. They also found that investors who had a higher IQ and were more successful produced higher Sharpe ratios, traded less frequently, held shares for longer periods of time, and maintained portfolios with larger numbers of different shares (implying greater portfolio diversification).

Perhaps some individual investors—for example, the buy-and-hold types with diverse portfolios composed of "value stocks"—are more rational than others such as the typical day trader. Is the seemingly persistent self-defeating behavior of the typical day trader consistent with rational investing? Why should we discuss rational investors? Might rationality or lack thereof explain why some investors outperform others?

Many financial models assume that all investors and all corporate managers are rational individuals who prefer more money to less and seek to maximize their wealth. Rational investors process information efficiently and use it to maximize their wealth. But, do investors really behave this way? Should they? The theories and models of efficient capital markets all draw from the assumption that investors are greedy, rational, and can access and exploit new information efficiently. By greedy, we mean that investors will strive to maximize their wealth. By rational, we mean that investors efficiently process and use all available information to maximize their wealth. Such assumptions might be quite

Financial Trading and Investing
DOI: https://doi.org/10.1016/B978-0-323-90955-6.00011-2

© 2023 Elsevier Inc. All rights reserved.

reasonable—at least to an extent, and allow for simpler and more straightforward model building. The efficient market theories are attractive, in part, because they facilitate financial model building. Financial model-builders can treat investors as participants in a mathematical game if investors' behavior is rational. Models based on rational behavior are usually less contrived, easier to create, and more believable than theories based on irrational behavior. Rational investors should maintain better survival rates in the marketplace, control more dominant wealth levels, and be more influential in price setting.

On the other hand, ask yourself how many investors you know that are truly rational. As reasonable as the rationality assumption might seem, and as compelling as the models that it leads to can be, does the rationality assumption really reflect the mind of the investor? This chapter explores the extent to which investors might or might not be rational and how such deviations from rationality might affect financial markets. Experiments by psychologists and perhaps our own observations of market behavior suggest that the "rational man" assumption is not fully justified. Behavioral finance is concerned with the actual behavior and thinking of individuals who make financial decisions.

The St. Petersburg Paradox and the Expected Utility Paradigm

In 1713, the mathematician Nicholas Bernoulli reasoned that a rational gambler should be willing to buy a gamble for its expected value. For example, it seemed rational for a gambler to invest up to $1 for a gamble that paid either $2 or zero based on the toss of a coin. He extended his reasoning to a series of coin tosses, continuing to reason that the ultimate value of the more complex gamble should still be its expected value. His cousin, Daniel Bernoulli, presented his paradigm in 1738 at a conference of mathematicians in St. Petersburg.[1] His extended problem, commonly referred to as the St. Petersburg Paradox, was concerned with why gamblers would pay only a finite sum for a gamble with an infinite expected value. Suppose, in Bernoulli's paradigm, the coin lands on its head on the first toss, the gamble payoff is $2. If the coin lands tails, it is tossed again. If the coin lands heads on this second toss, the payoff is $4; otherwise, it is tossed a third time. If the coin lands heads on the third toss, the payoff doubles again to $8; otherwise, it is tossed again for a potential payoff of $16. The process continues until the payoff is determined by the coin finally landing heads. Where n equals infinity, the expected value of this gamble is determined by the following equation:

$$E[V] = (0.5^1 \times 2^1) + (0.5^2 \times 2^2) + (0.5^3 \times 2^3) + \cdots + (0.5^\infty \times 2^\infty)$$

This equation is based on the expectation that the probability of the coin landing heads on the first (or any) toss equals 0.5. If the coin lands heads on the first toss, the payoff equals $\$2 = 2^1$. Since there is a 50% chance that the coin will land tails on the first toss and a 50%

[1]Nicholas Bernoulli first proposed this problem in a letter to Pierre Raymond de Montmort dated September 9, 1713, who then published it in his book later that year. The Swiss mathematician Gabriel Cramer actually proposed an essentially identical solution 10 years before Daniel. Correspondence between Nicholas Bernoulli and Cramer is available at http://www.cs.xu.edu/math/Sources/Montmort/stpetersburg.pdf#search=%22Nicolas%20Bernoulli%22. Daniel further argued in his essay that risk-averse investors should diversify, a rather novel concept at the time.

chance the coin will land heads on the second toss, the probability of achieving a payoff of $\$4 = 2^2$ on the second toss is $0.5 \times 0.5 = 0.5^2$. Thus, the payoff 2^n is realized with probability equal to 0.5^n. The expected value of the gamble equals the sum of all potential payoffs times their associated probabilities. So, exactly what is the expected value of this gamble? We simplify the equation above as follows:

$$E[V] = (0.5^1 \times 2^1) + (0.5^2 \times 2^2) + (0.5^3 \times 2^3) + \cdots + (0.5^n \times 2^n)$$
$$E[V] = (0.5 \times 2) + (0.5 \times 2) + (0.5 \times 2) + \cdots + (0.5 \times 2)$$
$$E[V] = (1) + (1) + (1) + \cdots + (1)$$

It appears, since there is some possibility that the coin is tossed tails an infinity of times ($n = \infty$), the expected or actuarial value of this gamble is infinite. This seems quite obvious from a mathematics perspective. Paradoxically, Bernoulli found that none of the esteemed mathematicians at the conference would be willing to pay an infinite sum (or, in most cases, even a large sum) of money for the gamble with infinite actuarial value. Were the mathematicians simply irrational? Or, should the worth or market value of a gamble or investment be less than its actuarial or expected value?

Bernoulli opined that the resolution to this paradox is the now commonly accepted notion of "diminishing marginal utility," which holds that as a person's wealth increases, the satisfaction that she derives increases, but at a lesser rate (see Figure 11.1). More money produces more satisfaction, but the rates of increase in satisfaction are less than the rates of increase in wealth. So the worth of a gamble to an investor is less than its expected value because the utility derived from each dollar of potential gains is less than the utility of each dollar potentially lost. Bernoulli proposed a log-utility function where an individual's level of satisfaction derived from wealth is related to the log of her wealth level. The key to this utility function is that satisfaction increases as wealth increases, but at decreasing rates. This means that an investor stands to lose more satisfaction in an actuarially fair gamble than she stands to gain. The potential loss in a "double-or-nothing" bet is

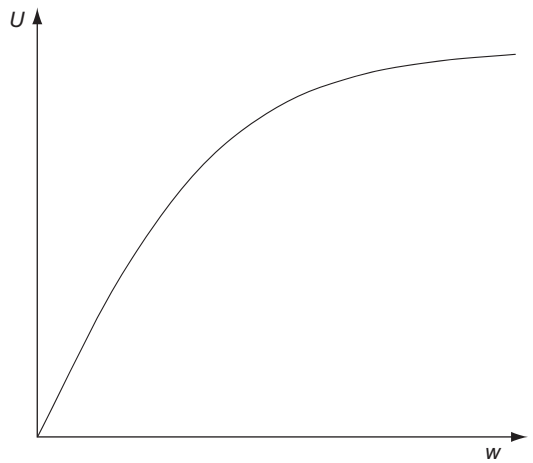


FIGURE 11.1 Utility of wealth $U = ln(w)$.

more significant than the potential gain. Thus, investors will reject actuarially fair gambles because, on average, they lose satisfaction or utility.

The implication of the utility function is that rational investors should seek to maximize the expected utility of their wealth, not expected wealth itself. Furthermore, this theory of utility can serve as the theoretical foundation for risk aversion. Thus, rational investors can be motivated not only by greed, but by fear as well.

Von Neuman and Morgenstern: Axioms of Choice

In their seminal treatise on game theory, von Neumann and Morgenstern (1947) present a set of behavioral assumptions (axioms) that we will adapt and use to derive the expected utility hypothesis. We shall start by assuming that investors make investment selection x_i from a convex subset $\mathbf{x}$ of the n-dimensional Euclidean space $\mathbb{R}^n$. The element x_i represents the number of units of investment i that can be selected by the investor from the n elements (investment combinations) of vector $\mathbf{x} = [x_1, x_2, \ldots x_n]^T$. Convexity of this subset means that the investor can form linear combinations any elements in $\mathbf{x}$ to obtain new elements in $\mathbf{x}$; that is, for any $\alpha \in [0,1]$, $\alpha x_j + (1 - \alpha)x_k \in \mathbf{x}$. The first three axioms ensure investor rationality:

1. *Reflexivity*: For an entire set $\mathbf{x}$ of investment alternatives $x_i, x_j \gtrsim x_j$ (x_j is at least as desirable as x_j). This axiom might be regarded as merely a mathematical necessity. This formal axiom is of little intuitive importance for our purposes.
2. *Completeness (or Comparability)*: For an entire set of investment alternatives x_i, either $x_j \succ x_k$ (x_j is preferred to x_k), $x_j \prec x_k$ (x_j is less desirable than x_k) or $x_j \sim x_k$ (x_j is equally desirable to x_k) for all j and k. Thus, the investor can fully specify his preferences over the entire set of investments.
3. *Transitivity*: For any x_i, x_j, x_k, if $x_i \succ x_j$ and $x_j \succ x_k$, then $x_i \succ x_k$. This axiom ensures consistency among choices.

While the three axioms listed above are sufficient to ensure investor rationality, working with such preference relations is difficult at best when n is very large. Hence, it is useful to develop and apply a rule that assigns values to choices. Such a rule is known as a cardinal utility function. A cardinal utility function assigns a unique number (utility level) to each and every choice (e.g., wealth level, risk/return combination, etc.). Three more axioms are needed to establish a cardinal utility function:

1. *Strong Independence*: If $x_i \sim x_k$, then for any $\alpha \in [0,1]$, $\alpha x_i + (1 - \alpha)x_j \sim \alpha x_k + (1 - \alpha)x_j$. It may be useful to interpret α as a probability. This axiom implies that preference rankings are not affected by inclusion in more complicated arrangements.
2. *Measurability (or Intermediate Value)*: If $x_i \succ x_j \succ x_k$, then there exists some α such that $\alpha x_i + (1 - \alpha)x_k \sim x_j$. This implies nonexistence of lexicographic (dictionary) orderings. Lexicographic orderings imply discontinuities in utility functions.
3. *Ranking*: Assume that $x_i \succ x_j \succ x_k$ and $x_i \succ x_m \succ x_k$, and $x_j \sim \alpha x_i + (1 - \alpha)x_k$ and $x_m \sim \gamma\, x_i + (1 - \gamma)x_k$ where $\gamma \in [0,1]$. Then it follows that if $\alpha \succ \gamma$, $x_j \succ x_m$ or if $\alpha = \gamma$, $x_j \sim x_m$.

These six axioms are sufficient to construct a cardinal utility function where utility can be represented with numbers. We shall usually add two more assumptions to this list:

1. Greed: Investors prefer more to less wealth.
2. Risk aversion: Investors prefer certainty to uncertainty. (This assumption need not always apply, but it does seem realistic.)

While seemingly complicated, these axioms of choice are the basis for the microeconomics of investment theory. Positive economics is typically based on the assumption that investors behave consistently with these axioms so as to maximize the utility associated with expected wealth. Perhaps more importantly, these axioms illustrate the mindset of economists analyzing the behavior of investors. The question we pose in the next section is whether and the actual extent to which investor behavior is consistent with these axioms.

11.2 PROSPECT THEORY

In the late 1960s, a cognitive psychologist named Amos Tversky began to find evidence that investors and gamblers might not behave in a manner consistent with maximizing utility. Tversky, in collaboration with others, including Daniel Kahneman (Nobel Prize winner for this body of work), had conducted numerous experiments on subjects testing decision-making characteristics. Generally, these experiments suggest that most subjects are quite irrational and inconsistent in their decision making. They have also written on applications of their results to financial decision making, suggesting that investors may actually be quite different from the theoretical "rational economic man." Kahneman and Tversky (1979) were particularly interested in how people made decisions when faced with uncertainty calling their new perspective *prospect theory* (with no particularly good reason). While expected utility maximization focuses on levels of wealth, prospect theory focuses on changes in wealth. Consider, for example, the following very simple illustration. Investor A has $1,000,000 right now, having doubled his money over the past 24 hours. Investor B has $1,500,000 right now, having lost 50% of his wealth over the past 24 hours. Which investor is happier now? Is more wealth always preferred to less? Do people make decisions as though this is true? Now, consider some slightly more complex scenarios.

Losses and Inconsistency

Consider the following example choice of gambles:

Gamble A: 0.33 probability of receiving 2500, 0.66 of receiving 2400, and 0.01 of receiving 0
Gamble B: 100% probability of receiving 2400

Kahneman and Tversky found that 82% of their experiment participants preferred Gamble B to Gamble A. However, they offered the same set of participants the following second set of gambles:

Gamble A*: 0.33 probability of receiving 2500, 0.67 of receiving 0
Gamble B*: 0.34 probability of receiving 2400 and 0.66 of receiving 0

In the second part of this experiment, 83% of participants preferred Gamble A* to B*. The same change was made to both gambles in moving from the first to second sets; 0.66 probability was shifted from Gambles A and B to A* and B* from winnings of 2400 to zero. The gamble shifts were identical, but many participants reversed their preferences. Yet from the first to second sets of choices, the changes to both gambles were identical; losses of 2400 were imposed on both gambles from the first set to the second set with probability 0.34.[2] Since the losses were identical, participants should not have reversed their decisions, but, clearly, the majority did. Kahneman and Tversky surmised that people are risk averse when evaluating positive outcomes (winnings), but risk-seeking when evaluating losses. Hence, people have diminishing utility of wealth functions with respect to winnings, but increasing marginal utility when faced with wealth decreases. Investors seem to exhibit similar reactions to reductions in wealth.

Consider a very simple variation on this problem. One group of subjects was presented with this problem:

1. In addition to whatever you own, you have been given $1000. You are now asked to choose between:
   A: A sure gain of $500
   B: A 50% chance to gain $1000 and a 50% chance to gain nothing
   A second group of subjects was presented with another problem.
2. In addition to whatever you own, you have been given $2000. You are now asked to choose between:
   A*: A sure loss of $500.
   B*: A 50% chance to lose $1000 and a 50% chance to lose nothing.

In the first group 84% chose A. In the second group 69% chose B*. The two problems are identical in terms of terminal wealth to the subject; each offers either $1500 with certainty or a 50-50 chance at $1000 or $2000. However, the phrasing of the question causes the problems that invoke different emotional responses. This leads to the following framing versus substance problem.

Frames versus Substance

Determining the preferences of individuals is complicated by their apparently inconsistent responses to questions that are presented to them. These inconsistencies not only make it more difficult to study investor behavior, they make the problem of offering financial advice rather problematic. For example, Kahneman and Tversky (1981) and Tversky and Kahneman (1986) describe how various individuals form decisions when identical questions are phrased differently. In particular, consider the following example when individuals are asked from two different perspectives to select from radiation or surgery for cancer treatment:

Survival Frame
Surgery: Of 100 people having surgery, 90 live through the postoperative period, 68 are alive at the end of the first year, and 34 are alive at the end of five years.

[2]See Exercise 6 at the end of this chapter for a more detailed discussion concerning the equivalence of these two gambles.

Radiation: Of 100 people having radiation therapy, all live through the treatment, 77 are alive at the end of the first year, and 22 are alive at the end of five years.
Mortality Frame
Surgery: Of 100 people having surgery, 10 die during surgery or the postoperative period, 32 die by the end of the first year, and 66 die by the end of five years.
Radiation: Of 100 people having radiation therapy, none die during treatment, 23 die by the end of one year, and 78 die by the end of five years.

Although the information presented in the "survival frame" is identical to that presented in the "mortality frame," 18% of respondents preferred radiation therapy when presented with the "survival frame," compared with 44% when presented with the "mortality frame." Interestingly, physicians and statistically trained business students did not react differently from other respondents when surveyed.

Consider an even simpler example suggested by Ritter (2003). How would you feel about paying a "peak-period surcharge" to dine in a restaurant? Next, consider how this surcharge pricing scheme differs from an "early bird" or "after-theater" discount. Presentation does matter. In any case, not only do we have evidence that individual decision making is inconsistent, but preferences may be inconsistent and individuals irrational as well. Consider this problem in an investments context. How can an investment advisor assess her client's level of risk aversion when the framing of the question affects the investor's response? Can we really expect investors to behave consistently when their risk attitudes change depending how their situations are addressed?

The framing issue presents itself regularly in promotional literature distributed by investment advisors, mutual funds, and other financial institutions, who regularly exploit investors' inability to contend with fallacious arguments. Consider the following description concerning corporate annual reports by Beattie and Jones (2000), "measurement distortion can occur through graphical devices such as a nonzero axis or a broken axis, which cause the rate of change in trend lines to appear greater than is actually warranted." Their paper argues that graphs are frequently used selectively and graph formats exploit measurement distortions, implying inaccurate depictions of financial performance. Diacon and Hasseldine (2007) find evidence that investors are regularly misled by such distorted displays when selecting investment funds.

There are a number of ways that the framing effect can surface in the investment decision-making process. Consider, for example, the return and risk effects of a decision to add a single stock to a portfolio. Investors often make decisions on a single stock based on its return/risk characteristics rather than how the stock's purchase will affect the return/risk characteristics of the investor's overall portfolio. This difference has the potential to be crucial, particularly if the investor's portfolio is or will become poorly diversified. A high-risk stock can be a very good contribution to a poorly diversified portfolio if the stock serves to decrease the risk of the portfolio. In related decision-making arenas, pension, insurance, lump sums, and annuity decisions can be very confusing to consumers and investors who easily fall victim to framing effects (e.g., Brown, Kapteyn, Erzo, Luttmer, & Mitchell, 2017).

Maintaining the Status Quo: Joe and His Opera Tickets

The following story was taken from the *Wall Street Journal*[3]:

> On the way to the opera, Joe loses his pair of $50 tickets. Most likely, he will not buy another pair—spending [a total of] $200 [including $100 on the lost tickets] to hear "La Boheme" seems a bit much. But suppose, instead, he arrives at the theater tickets-in-hand, but discovers he has lost a $100 bill. He could sell his tickets, which would net him the same result as in the first case—out $100 and out the tickets. But he probably won't sell.... Joe may think he is entirely rational, but he leans consistently toward the status quo.

This particular type of bias to maintain the status quo is sometimes referred to as the *endowment effect*. This effect causes losses or what is given up to weigh more heavily in the decision-making process than gains or what is acquired. This effect seems to manifest itself in investing through a seeming reluctance to sell stocks, particularly stocks that have lost value. Numerous studies (e.g., Shefrin & Statman, 1985) have documented investors' reluctance to sell their "losers" to capture tax write-offs. The tax write-off implications of selling a stock that has lost value are enhanced when "losers" are sold before year's end. Tax loss write-offs captured on that year's return filings have more value than delayed loss write-offs. However, most investors demonstrate a pronounced reluctance to sell their losers to capture their write-offs. Investors seem intent on maintaining their status quo even when changing their portfolios would result in increased wealth. Some observers refer to this phenomenon as "fear of regret." More generally, studies have suggested that this endowment effect or disposition effect might lead stock markets to underreact to news (Frazzini, 2006) and exacerbate momentum effects in stock prices (Grinblatt & Han, 2005).

Anchoring

A similar sort of problem exists with "anchoring," where the decision maker places undue emphasis on some factor, number, or measure. For example, Kahneman and Tversky asked participants in an experiment to spin a roulette wheel with numbers from 1 to 100 and then estimate the number of countries in Africa. They found that participants' estimates were unduly influenced by the result of the roulette wheel spin result. Low roulette wheel outcomes were followed by lower estimates of the numbers of African countries. Similarly, Genesove and Mayer (2001) found that sellers of houses and apartments tend to be unduly influenced by purchase prices of their homes. There is similar evidence suggesting that investors may be unduly biased by purchase prices of their securities, and that such anchoring can lead to price momentum effects (e.g., Liao, Chou, & Chiu, 2013). Several studies such as Beattie and Jones (2000) have found that amateur traders are more affected by endowment and anchoring effects than professionals.

[3]This article entitled, "Intrinsic Value: Outsider who Challenged Dismal Science," by Roger Lowenstein in the June 6, 1996 issue of the *Wall Street Journal* provided a number of examples of the work of Amos Tversky, the Stanford psychologist who studied irrational consumer and investor behavior.

11.3 BEHAVIORAL FINANCE

Behavioral finance, largely rooted in prospect theory, is concerned with the impact of human emotions and cognitive impairments on investment decision making. Economists have used expected utility maximization for over two centuries as the presumed objective of the rational investor. Expected utility maximization has served and still serves as the "industry standard" for modeling investor behavior. In this section, we review a number of studies concerning individual and investor rationality. Many of these studies were conducted by experimental psychologists while most of the remainder are based on statistical analyses of financial data.

The case for assuming capital markets efficiency when developing financial models is strong. Models based on rationality tend to be simple and one can argue that rational investors produce most market prices despite a large contingent of irrational investors. However, regardless of their intuitive appeal, theoretical models should not be accepted without consideration for empirical evidence. There exists much controversy in academia and in the profession regarding the extent to which models of market efficiency reflect reality. The following sections and the next chapter represent a sampling of this literature.

The Monty Hall Judgment Error

One very interesting experimental study of investor irrationality, failure to learn, and market rationality is presented in Kluger and Wyatt (2004). This problem is adapted from the well-known Monty Hall problem, based on the late 1960s game show "Let's Make a Deal."[4] In the show, Monty Hall, its host, would offer contestants an opportunity to choose one prize hidden behind one of three identical doors. Prizes hidden behind two of three doors were worthless (if the contestant selected either of these doors, she was "zonked") but the prize hidden behind the third was valuable. The contestant would attempt to choose the door behind which the valuable prize was to be hidden. At this stage in the game, the contestant had a one-third probability of winning the valuable prize. Before allowing the contestant to see whether she had won the valuable prize, and with increasing audience anticipation, Mr. Hall, who knows what is behind the doors, would then typically show the contestant the worthless prize behind one of the two doors that the contestant did not select. He would then offer the contestant an opportunity to switch her selection to the prize behind the third door. Thus, the contestant's problem is whether to stick with her original selection or to switch her selection to the prize hidden behind the third door.

To summarize, the valuable prize is behind one of the three doors. The probability of the contestant having selected the door with the valuable prize was initially 1/3. But, once one of the two worthless prizes is revealed behind a second door, can the contestant improve his odds of obtaining a valuable prize by switching his decision to select the other unopened door? Think about this question before reading on.

[4]Monty Hall was the stage name for Monte Halparin, game show emcee, producer and singer.

Regardless of what prize remains behind the first door selected by the contestant, Mr. Hall will reveal the worthless prize behind a second door. Hence, the probability of a valuable prize behind the first door remains 1/3. We know that Mr. Hall will not reveal the prize behind the originally selected door, so its probability of being the desirable prize is unchanged. The door that Mr. Hall will select to open will have a worthless prize with probability one. This door will not have a valuable prize behind it. What is the probability that the valuable prize is behind the third door? This probability must be $1 - 1/3 - 0 = 2/3$. Why? Remember that Mr. Hall will not open a second door with a valuable prize behind it. This doubles the probability that the valuable prize is behind the third door. Hence, the contestant should always switch his selection to maximize his probability of obtaining the valuable prize.[5] Most contestants did not. In fact, most experimental subjects, when provided a description of this problem announce that they would not switch (Friedman, 1998). When asked to calculate probabilities, many people will say that the valuable prize is equally likely to be behind the first and third doors, despite the additional information realized when Mr. Hall reveals the prize behind the second door.

Most people have no difficulty estimating that the initial probability of 1/3 for the prize behind any one of the three doors. This heuristic has served most people well for years. However, people tend to use the same heuristic when a "zonk" is revealed behind one of the doors, leading them to conclude that there is a 50/50 probability that the prize is behind one of them. This heuristic is difficult to abandon when the nature of the problem changed, as the problem solution shifts from an unconditional probability regime to a less intuitive conditional probability. Most people refuse to accept this shift.

Perhaps, more interestingly, most subjects refuse to accept the validity of mathematical proofs offered to demonstrate the wisdom of switching doors. Furthermore, most subjects continue to refuse to switch doors after being permitted to watch repeated trials of this experiment where the third door leads to the valuable prize with a frequency of approximately 2/3. Hence, subjects seem to either never learn from their errors or learn very slowly.

Kluger and Wyatt (2004) conducted experiments to determine how a market might behave in such a scenario. Kluger and Wyatt gathered subjects in a laboratory setting and had them compete to select investments whose payoffs were "behind doors." "Investors" participated in repeated trials, were offered opportunities to select doors, and then

[5]If you are not convinced that this solution to the Monty Hall problem is correct, consider the following table. The contestant has three choices to start with, followed by two choices after each of the first three. Thus, there are six possible outcomes, including three that result from switching doors. In two of the three switching outcomes, the contestant gets the car, compared to only one out of the three if the contestant keeps his original door selection.

Choose a door with a car behind it		**Choose a door with a goat behind it**		**Choose a door with a goat behind it**	
Keep Door	Switch Door	Keep Door	Switch Door	Keep Door	Switch Door
Get a Car	**Get a Goat**	Get a Goat	**Get a Car**	Get a Goat	**Get a Car**

compete to pay to either retain or switch their selections. If investors were rational and understood the nature of the Monty Hall problem, market prices to switch should be twice the levels of prices to retain original selections. Kluger and Wyatt found that when all investors mispriced the selections, prices of original selections remained roughly comparable to prices to switch. Hence, misinformed or irrational investors consistently mispriced the investments. However, when as few as two "rational investors" who correctly estimated the probabilities were included in the trials, prices to switch were roughly double the prices to retain original selections. Hence, it seemed that competition between only two rational investors out of many were necessary for market prices to reflect rational probabilities. This experiment raises an interesting question: Can as few as two competing rational investors offset irrationality in prices in an otherwise irrational securities market?

Similarly, traders all have good reason to question their abilities to beat the market. First, at least 50% of traders are likely to be "below average," and underperform the market, even without transaction costs. Transaction costs worsen traders' abilities to outperform random "buy-and-hold" strategies. The market efficiency hypotheses provide ample rationale to avoid the attempt to "out trade" the market. Statistical evidence demonstrating the general inability to outperform the market is overwhelming.

Dumb, Dumber, and Dead

The Monty Hall illustration result is quite counterintuitive and has caused confusion even among trained mathematicians. However, investors are susceptible to even more basic errors. For example, Rashes (2001) documents numerous cases of strong correlations between stocks with similar ticker symbols. For example, shares of stock for Massmutual Corporate Investors (ticker: MCI), an NYSE listed fund that did not hold shares of any major telecommunications companies, were strongly correlated with those of MCI Communications (ticker: MCIC), realizing particularly high volume and price during the period of merger discussions and activity involving the telecommunications firm. Massmutual's returns were far more correlated with MCI's than with AT&T or any of the other telecommunications firms. In fact, there is evidence that investors bought shares of Massmutual and held them for long periods of time, believing that they had invested in MCI. Similarly, the most volatile trading day for the Castle Convertible Fund (ticker: CVF) was April 15, 1997, after an article appeared in the *Financial Times* about investments in fraudulent companies by the Czech Value Fund (abbreviated as CVF in the article). In a 22-minute span, the stock price dropped from 24.75 to 16.75 before closing the day's session at 23. On June 24, 1998, an AT&T bid for shares of Tele-Communications Inc. led to a 4.3% increase in the price of Transcontinental Reality Investors Inc. (ticker: TCI). In fact, the highest trading volume ever experienced by Transcontinental Reality shares to that date was on October 13, 1993, when Bell Atlantic announced its intent to purchase Tele-Communications.

Kadapakkam and Zhang (2014) examine shares of bankrupt firms after courts confirm that shareholders will receive nothing, finding that some have substantial trading volume at positive prices despite being intrinsically worthless. For example, Motors Liquidation Company (the former pre-bankruptcy GM) had daily trading volume peaking at over

$250 million and BB Liquidation (formerly Blockbuster) averaged 1.7 million shares while awaiting their cancellation. Covering short sales did not explain this trading volume, though short-selling did help mitigate irrational price increases. "Fake news" on various Internet sites seemed to help prop up share prices and perhaps support "pump and dump" schemes.

Even options markets exhibit evidence of uninformed or perhaps even irrational behavior. Studies have found that more than half of outstanding long-call-option positions that should have been exercised on the day before the ex-dividend date went unexercised, and many deep in-the-money puts that should have been exercised early were not.[6] While approximately half of this failure to exercise can be attributed to transaction costs, the remainder seems to be due to lack of awareness of the early exercise decision, inability to monitor their option positions, or irrationality. Furthermore, certain early exercises of CBOE call options seem to be irrational, where customers of both full-service and discount brokers seem to exhibit irrational exercise behavior, while traders in large investment houses did not.

Perhaps the most notorious of underperforming funds during the 1990s bull market was the Steadman Technology & Growth Fund. Its returns during part of this decade were −5% in 1992, −8% in 1993, −37% in 1994, −28% in 1995, −30% in 1996, and −28% in 1997. Market returns were positive in each of these years. Only one of the four Steadman funds had a positive long-term return, although it was quite small. Why were returns so dismal? Portfolio turnover rates and transaction costs were extremely high, expense ratios were frequently in the 6% to 7% range, and Steadman seemed not to have a coherent investment strategy, indiscriminately investing in stocks and other securities. The SEC forced the fund to stop accepting money from new investors in 1989. Why did other investors keep money invested throughout the dismal performance of the early 1990s? When the Technology & Growth Fund was finally shut down, about 30% of the redemption checks were returned, presumably because the shareholders were dead.

Myopia and Overreaction

Several studies have suggested that investors overreact to the most recent information (e.g., Thaler, Tversky, Kahneman, & Schwartz, 1997), and that they are too influenced by short-term price swings than they should be. In addition, Griffin and Tversky (1992) and Sorescu and Subramanyam (2006) find that investors tend to overreact to dramatic upgrade recommendations of less experienced and less reputable analysts while they under-react to less dramatic upgrade recommendations of more experienced and more reputable analysts. Furthermore, there is some empirical evidence (to be detailed later) that stocks experiencing wide short-term price swings reverse those swings in the intermediate term. This behavior suggests that investors may overreact to information. In addition, there is much evidence in the psychology and investments literature that people tend to be overconfident in their own judgments, perhaps, oddly enough, even enabling certain overconfident traders to generate higher profits than their more rational competitors under

[6]See Pool, Stoll, and Whaley (2008) and Barraclough and Whaley (2012) for failure to exercise early and Poteshman and Serbin (2003) for irrational early exercise.

limited conditions. While, in general, overconfident investors seem to trade more frequently and realize reduced returns, in more professional trading settings, more confident traders may realize higher profits by more frequently exploiting bid–offer spreads. There is also evidence that experts tend to be more prone to overconfidence than novices. Overconfident professional traders also tend to be more aggressive in their trading strategies, and in some instances, improving their returns.

Myopia

Consider a paper by Benartzi and Thaler (1995) that examines the widespread large differences between bond and stock returns. From 1925 to 1997, intermediate term government bond returns have averaged approximately 5.2% while stock returns have averaged approximately 10.7%. This spread indicates reluctance on the part of investors to buy stock, even though for practically every intermediate- to long-term period over the 75 years prior to 1997, investors could have expected to earn more in stock markets.[7] Benartzi and Thaler argue that even long-term investors have a myopic outlook, causing them to make every effort to avoid short-term losses and risks. This aversion to losses is often referred to as *myopic loss aversion* and is frequently offered as the cause for the equity return premium relative to bonds. For example, their experiments conducted on university employees indicated that when shown only the 30-year returns on stocks and bonds, employees preferred to hold more stock in their pension accounts. Thus, when employees see how much money will be available in their retirement accounts in 30 years, they prefer the higher returns associated with stocks. However, when shown additional information reflecting the entire year-by-year return path over the 30-year period, employees preferred less stock in their accounts, even though they still would have more money at the end of the period with stock investment. Thus, investors with a long-run orientation to investment still maintain a myopic aversion to risk because they care not only about how much money they will have available when they retire, but their gains and losses en route to their retirements. A number of financial institutions, such as Bank Hapoalim, the largest manager of mutual funds in Israel, reduced the frequency with which they issue financial statements to their investors, citing investors' myopic risk aversion as their rationale (see Gneezy, Kapteyn, & Potters, 2003). They expected that their investor-clients would be less likely to move money out of their funds if they received reports of balances less frequently.

Jeremy Siegel, in his book titled *Stocks for the Long Run* (1994), lends support for this myopic loss aversion argument, arguing that, "It is widely known that stock returns on average, exceed bond returns in the long run. But it is little known that in the long run, the risks in stocks are less than those found in bonds or even bills!" While this statement might overstate the risk/return point, it does seem that most investors overemphasize short-term risk at the expense of long-term returns. On the other hand, Dimson, Marsh, and Staunton (2002) note that the long run can be very long. For example, the Italian stock market failed to keep pace with inflation from 1915 to 1978 and

[7]Note that their study period did not include the slumping stock market period 2000 to 2009. The same is true of Siegel's study, discussed in the next paragraph.

communist takeovers closed down Russian and Chinese stock markets in the 20th century.

Numerous experiments reported in the psychology literature have indicated that individuals in their decision making tend to overemphasize recent information and trends and underemphasize prior information. For example, Shiller, Kon-Ya, and Tsutsui (1996) found that at the peak of the Japanese market, 14% of Japanese investors expected a crash. However, just after the market did crash, 32% expected another crash. Even professional analysts tend to overreact to new information. Excessive dependence on recent information may lead the market to overreact to new information, and may even explain the so-called price-to-earnings ratio anomaly (discussed in Chapter 11). In an important paper, Shiller (1981) examines empirical data and argues that the volatility in stock markets cannot be explained by the volatility of cash flows (dividends) associated with stocks, suggesting that investors may overreact to risk. Market prices are extremely sensitive to potential future earnings anticipation, and in fact, are too sensitive, so that they vary far more than the actual dividends.

Hyperbolic Discounting

Suppose that you just won the lottery. Would you rather be paid $500,000 in five years or $1,000,000 in ten years? Many, but not all people would opt for the larger sum delayed by an additional five years. Now, change the payout dates. Would you rather have $500,000 now or $1,000,000 in five years? Many more people would opt for the smaller sum now. That is, at least a few people would reverse their preferences by advancing the payoff dates by five years. Similarly, as cash flow dates approach, some investors will reverse their preferences toward preferring smaller cash flows sooner.

The time value principle holds that preferences should not depend on the decision date. That is, if an investor prefers Investment A to B today based on its current time value of money, she would still prefer A to B at a later date given that the cash flow dates do not change. For example, suppose that an investor (consumer) prefers alternative A to alternative B (A > B, or $CF_A e^{-rT_A} > CF_B e^{-rT_B}$ with $r > 0$, where alternative A has a higher nominal value or cash flow ($CF_A > CF_B$), but is delayed longer ($T_A > T_B$). Thus, the investor (consumer) decides now that she prefers to consume the larger amount CF_A at the later date T_A to the smaller CF_B sooner at date T_B. However, if both alternative investment payoff levels were advanced by the same increments in time T_B (e.g., alternative A now pays at time ($T_A - T_B$) and B now pays at time $T_B - T_B$ or time zero – now), many decision makers would reverse their preferences. That is, even though they might initially prefer the larger awards even if delayed longer, if the smaller reward were offered now, they would prefer it to the larger reward with a shorter delay. That is:

$$CF_A e^{-rT_A} > CF_B e^{-rT_B}$$

$$CF_A e^{-r(T_A - T_B)} < CF_B e^{-r(T_B - T_B)}$$

These two statements are mutually inconsistent, and illustrate temporal myopia. Consider the following example, with cash flow present values set forth in Table 11.1. An investor has the opportunity to decide between receiving $50 five years from now and $100 ten years

TABLE 11.1 Hyperbolic Discounting

t	CF_A	CF_B	$PV(CF_A)$, $r = 0.1$	$PV(CF_B)$, $r = 0.1$	$PV(CF_A)$, Hyperbolic Discounting	$PV(CF_B)$, Hyperbolic Discounting
0			36.7879445	30.32653314	36.78647301	24.28315774
1			40.65696636	33.51600244	38.88817407	25.67051386
2			44.93289679	37.04091115	41.10995044	27.13713301
3			49.65853074	40.93653774	43.45866232	28.6875437
4			54.88116395	45.24187095	45.94156184	30.32653314
5		50	60.65306629	50	48.56631548	50
6			67.03200488		51.34102771	
7			74.0818223		54.27426602	
8			81.87307548		57.37508741	
9			90.4837419		60.65306629	
10	100		100		100	

from now. If she discounted cash flows at an annual rate of 10% ($PV = CFe^{-0.10}T$), she would prefer to have \$100 ten years from now. Suppose that the investor had "impulse issues," and that her discount rate were 50% for one year immediately prior to consumption and 5.556% for all other years: $PV = CFe^{-(0.05556+\cdots+0.05556+0.5)}$. Still, the investor would prefer to receive \$100 in 10 years to receiving \$50 in five years. However, if the investor were permitted to make her decision in the fifth year, her preference ranking would reverse: $PV_A = 48.47 < PV_B = 50$.

This investor uses a form of hyperbolic discounting to discount her cash flows. The discounting functions are based on a constant discount rate $r = 0.10$ from above and can be easily generalized to accommodate time-varying discount rates:

$$CF_A e^{-(r_1+r_2+\cdots+rT_B+\cdots+rT_A)} > CF_B e^{-(r_1+r_2+\cdots+rT_B)}$$

$$CF_A e^{-(r_1+\cdots+r_{(T_A-T_B)})} < CF_B e^{-(r_{(T_B-T_B)})} = CF_B$$

which is still inconsistent unless r_1 (along with, perhaps other $r_t < r_{TB}$) is sufficiently high. This type of hyperbolic discounting was proposed by Strotz (1956) and Laibson (1997).

Thus, Table 11.1 suggests that the consumer will prefer option A in every period except when the consumption of option B is imminent. The preference ranking changes because immediate consumption is considered (discounted) differently from deferred consumption. However, this impulse typically applies only when one of the options is immediate.

Neuroeconomics might offer an explanation for this effect.[8] Consider a scenario where experimental subjects in a laboratory experiment are offered a choice between \$10 now

[8]We will discuss neurofinance (which is concerned with the physiological effects of financial decision making) in greater detail shortly.

and $15 in one month. MRI analysis shows that both the separate limbic (which governs intuitive, affective, and emotional cognition) and analytic (which govern analytical, computational, and strategic cognition) systems show blood flow activity. Next, change the offer terms such that only two future award alternatives exist: $10 two weeks from now and $15 in one month. The MRI imaging changes such that only the analytic system shows activity. Subjects are more willing to accept the delayed, but larger rewards. Thus, when the choices include a current reward, the limbic system seems to allow the temptation of immediate gratification to dominate the otherwise apparently more rational decision (Eat, drink and be merry, for tomorrow, we may die). Hence, precommitment (policy, monetary deposits, contracts, etc.) to more rational decisions might help overcome this effect. Similarly, reliance on others for decision making may also be useful when their analytical systems are more involved in this process.

Consider, for example, that Dr. Edward Miller, dean of the Johns Hopkins University Medical School and chief executive officer of its hospital, claims that over half a million people undergo coronary-artery bypass graft surgery every year in the United States, but only about 10% of these patients make lifestyle changes needed to prevent future surgeries, chest pains, and premature death. It is well known that employees regularly opt not to participate in employer retirement savings plans, even when employers match their contributions. One might argue that contractual precommitment or government intervention might play a role in regulating human behavior. For example, O'Donoghue and Rabin (2006) suggest that willpower weaknesses might be remedied by taxing potato chips and subsidizing carrots. The U.S. federal government provides for tax deferral subsidies to encourage employer-sponsored retirement accounts.

Overreaction

Stocks that experience wide short-term price swings may have a tendency to reverse those swings in the intermediate term. This behavior suggests that investors may overreact to information. As we will discuss later, DeBondt and Thaler (1985) argue that the market overreacts to information and that abnormal returns can be realized by buying losers and selling winners. Their study indicated that buying stocks that performed poorly in a prior three- to five-year period and selling those that performed well would have generated higher than normal returns in subsequent three- to five-year periods. Most significant mean reversion tendencies seem to hold for the month of January.

Investor Moods, the Weather, and Investment Returns

There is some evidence that investor moods might significantly affect market performance. For example, Kamstra, Kramer, and Levi (2003) found that seasonal disorder, a medical condition where the shorter days in fall and winter lead to depression for many people, is associated with reduced stock market returns after adjusting for a variety of other factors. They found that stock market returns are higher during the spring quarter than in the fall quarter. Seasonally related factors can be captured by examining cyclical ties in markets with varying latitudes, paying attention to variations in behavior near

solstices and equinoxes. Northern and southern hemisphere returns seem six months out of phase.

Several studies have found that weather might affect market returns. For example, Hirshleifer and Shumway (2003) found that cloud cover in the city of a country's major stock exchange is negatively correlated with daily stock index returns in 18 of 26 national exchanges for the period of 1982 to 1997. Stock market performance was simply worse on cloudy days. In New York City, there was a 24.8% annualized return for all days forecast to be perfectly sunny, and an 8.7% average return occurred on cloudy days. There seems to be evidence in the psychology literature indicating that sunshine improves investor moods, and may collectively increase investor willingness to accept risk. Limpaphayom, Locke, and Sarajoti (2007) obtained consistent results using wind and the traders on the Chicago Mercantile Exchange, with windy days seeming to increase bid-ask spreads.

However, despite these dramatic return differences, the evidence on the relationship between investor behavior and weather is not entirely clear. For example, Goetzmann and Zhu (2005) analyzed trading accounts of 79,995 investors from 1991 to 1996, finding that individual investors' trading behavior is no different on sunny days than on cloudy days. Nevertheless, market-maker behavior was significantly affected by cloud cover in one important respect: cloudy days were associated with wider bid-ask spreads, suggesting that investors (or market makers) were more risk averse on these days.

Edmans, García, and Norli (2007) find evidence of a strong link between soccer outcomes and mood, using international soccer results as their primary mood variable. They found that markets experienced significant decline after soccer losses, such as losses in the World Cup elimination stage leading to next-day abnormal stock returns of 0.49%. These loss effects were stronger in small stocks (more likely to be traded by individual investors) and for more important games. They also documented loss effects after international cricket, rugby, and basketball games. They controlled for effects of related business revenues resulting from contest outcomes. While sports losses were clearly related to stock returns, they did not find that sports wins had any significant effects on returns.

Scientific evidence is clear that lunar cycles are related to tides, animal behavior, and other natural phenomena. In related research drawing on inconsistent research results indicating that homicide rates, hospital admissions, and crisis incidents all peak in the days around full moons, Dichev and Janes (2001) find evidence that stock returns around new moons nearly double those around full moons. The effects arise in U.S. stock markets over 100 years and in other markets over the past 30 years. Are there good explanations for this? Should we suspect that eventually anything can be proved or disproved with careful statistical analysis? Kim (2016), with a particular focus on the methodologies conducted in the cloud cover studies by Hirshleifer and Shumway (2003), argues that many such results are spurious, the product of mining large data sets with faulty statistical methods.

Simplifying the Decision Process

Even if investors were able to efficiently and rationally process small information sets for simple decisions, are they capable of processing larger sets of information for more

complex decisions? Consider the following case illustrated by W. Bentley MacLeod and presented by Andrew Lo in TIAA-CREF's Investment Forum:

> [C]onsider the simple task of getting dressed in the morning: For a typical male wardrobe of 5 jackets, 10 pants, 20 ties, 10 shirts, 10 pairs of socks, 4 pairs of shoes and 5 belts, there are two million different combinations to evaluate, and if we allow one second to evaluate each outfit, it would take about 23 days to select the "best" outfit.. . . Yet we all seem to get dressed in just a few minutes—how?

Even fairly simple problems such as getting dressed require heuristics or rules of thumb. However, all too often, these rules of thumb simply do not work for more complex problems such as selecting stocks. In his paper on stock market noise, Black (1986) wrote:

> Because there is so much noise in the world, people adopt rules of thumb. They share their rules of thumb with each other, and very few people have enough experience with interpreting noisy evidence to see that the rules are too simple.

Nevertheless, rules of thumb are important. Without them, many people would be simply unable to invest (or, even get dressed in the morning).

Rational Investors and Diversification

Perhaps the single most important lesson from modern finance is the importance of diversification and its role in the management of risk. Investors can create more efficient portfolios by investing in broader arrays of investments. However, many studies including Goetzmann and Kumar (2008) suggest that investors do not diversify efficiently. For example, investors tend to concentrate their holdings in shares of their employers, in domestic shares, and even in the shares of companies located near their residences. While many investors certainly have lost substantial portfolio efficiency due to this sort of myopia, a smaller number of investors have been able to outperform the market by under diversifying. For example, Coval, Hirshleifer, and Shumway (2005) document strong persistence in the performance of trades made by skillful individual investors who seem able to exploit market inefficiencies and information advantages to earn abnormal profits. Ivković and Weisbenner (2008) corroborate this result, finding that households exhibit a strong preference for local investments. Ivković and Weisbenner demonstrate that individuals' investments in local stocks outperform their investments in nonlocal stocks. They find that this effect is more pronounced for their investments in smaller companies outside of the S&P 500. Similarly, Massa and Simonov (2006) found that Swedish investors also exhibited strong preferences for holding stocks to which they are geographically or professionally close. They also found that such apparent information advantages enabled these investors to earn excess returns on their more "familiar" investments. In fact, Ivković, Sialm, and Weisbenner (2005) found that investors holding smaller, less diverse portfolios and concentrating their holdings in local companies actually outperformed their more "efficient" counterparts who held more diverse portfolios. Coval and Moskowitz (1999, 2001) found similar results for professional fund managers: professionals tend to invest more heavily in shares of companies located nearby and earn higher returns on the holdings of these local companies.

11.4 BODY AND MIND OF THE TRADER

Overconfidence

How many investors believe that they are better than average traders? How many drivers think that they are better than average?[9] How many people think that they are dumb (less intelligent than average)? Half? Certainly, no one begins trading stocks thinking that he will perform worse than the market or the average trader.[10] Yet, it is likely that half of all traders are below average. But, how do typical investors actually fare? And how do they fare relative to professionals? Russo and Schoemaker (1992) wrote:

> Good decision-making requires more than knowledge of facts, concepts and relationships, it also requires metaknowledge—an understanding of the limits of our knowledge. Unfortunately, we tend to have a deeply rooted overconfidence in our beliefs and judgments.

To test for overconfidence and compare results across professions, Russo and Schoemaker created and administered a 10-question test similar to the Trivia Quiz provided in Table 11.2.[11] Before discussing the results of their study, take the test yourself. Since you will probably not know the exact answers for each of these questions, your objective is to guess by setting minimum and maximum bounds for each of the questions such that you are 80% confident that the actual answer will be within the 80% confidence interval that you set. Obviously, if your range is from 0 to infinity for each question, the correct answers will all fall within your confidence intervals. But, again, your goal is to answer only with 80% certainty, so narrow your ranges accordingly. Take this test before reading further.

The answers to this test are provided in Table 11.3. Since takers of this test were to establish 80% confidence intervals for each of the question estimates, each person taking this test should have gotten eight questions correct—at least on average. Answering a question correctly simply means that the actual answer will fall within the 80% confidence interval. Thus, correctly answering eight questions indicates correct "calibration" of the confidence ranges. More than eight correct may indicate underconfidence and excess conservatism; less than 80% correct indicates overconfidence. Thus, the real purpose of this test is to examine the test-taker's ability to calibrate her confidence range. The vast majority of test takers find that their answers fall within 80% confidence intervals far less than 80% of the time; in fact 20% to 50% scores are typical. Russo and Schoemaker administered a test similar to this to members of many professions, finding that most members of most professions exhibit significant overconfidence. In fact, securities analysts performed poorly relative to members of other professions, receiving scores of only 36% when asked to set

[9] Svenson (1981) found in his survey of 22-year-old drivers that 82% considered themselves to be among the 30% safest.

[10] For example, Ito (1990) finds in surveys of foreign exchange traders that most are more optimistic about how rate changes will affect them than they are about how rate changes will affect their competitors.

[11] The actual test administered by Russo and Schoemaker is over 25 years old and has been replaced here by another set of 10 questions. A few readers will have some difficulty converting figures either into or from the metric system; the reader should adjust confidence intervals for these difficulties.

TABLE 11.2 Trivia Quiz: Enter the Range (Minimum and Maximum) for Which You Are 80% Certain the Answer Lies Within

	Min	Max
1. What was the population of the city of San Diego according to the 2020 census?		
2. In what year did Michelangelo complete painting the Sistine Chapel ceiling?		
3. What was the maximum weight in pounds of President William Howard Taft (the heaviest U.S. president)?		
4. What is the height in feet (3.3 feet per meter) of Mount Kosciuszko, the highest peak in Australia?		
5. As of 2022, how many years have elapsed since the birth of the Carthaginian military commander, Hannibal?		
6. What is the maximum heart rate in beats per minute of the blue throated hummingbird?		
7. In nautical miles (1.1 miles or 1.8 km per nautical mile), what is the maximum flight range of a Boeing 747-8?		
8. Excluding amendments and signatures, how many words are in the U.S. Constitution?		
9. What is average distance in miles of the earth's orbit around the sun from the planet Saturn's orbit?		
10. What is the weight in grams of the Hope Diamond?		

TABLE 11.3 Trivia Quiz Answers

1. What was the population of the city of San Diego according to the 2020 census?	1,386,932
2. In what year did Michelangelo complete painting the Sistine Chapel ceiling?	1512
3. What was the maximum weight in pounds of President William Howard Taft?	355
4. What is the height in feet of Mount Kosciuszko, the highest peak in Australia?	7310
5. As of 2022, how many years have elapsed since the birth of the Carthaginian military commander, Hannibal?	2204
6. What is the maximum heart rate in beats per minute of the blue throated hummingbird?	1250
7. In nautical miles, what is the maximum flight range of a Boeing 747-8?	8000
8. Excluding amendments and signatures, how many words are in the U.S. Constitution?	4543
9. What is average distance in miles of the earth's orbit around the sun from the planet Saturn's orbit?	800,000,000
10. What is the weight in grams of the Hope Diamond?	9.1

90% confidence intervals. Test subjects in most professions simply fail consistently to calibrate their certainty levels.

To test whether experiment subjects function better with respect to their areas of expertise, they were administered similar types of tests focusing on issues more relevant to their

professions. The majority of professionals in most fields failed to perform much better. Interestingly, only two of the tested professions exhibited calibration skills failing to exhibit overconfidence: accountants on tests related to auditing and climatologists when forecasting weather.

What does this mean for stock market behavior? First, trading activity increases when traders are overconfident. Overconfident traders tend to be more aggressive in their trading strategies. Overconfident traders underreact to the information content of trades by rational traders, causing positive serial correlation in returns. Odean (1998a, 1998b) and Barber and Odean (2000, 2001a, 2001b), in studies of trading in 10,000 and over 60,000 discount-brokerage accounts from 1987 to 1993 and from 1991 to 1996, found that trading by investors reduced their levels of wealth below what they would have realized with simple buy-and-hold strategies. For example, they found that stocks these investors sold beat the market while those they bought did worse than the market. By one year after the trades, the average investor ended up over 9% worse off than if had he done nothing. In another study of 1607 investors, Barber and Odean (2001a, 2001b) found that investors that had switched from phone-based trading to internet systems increased their portfolio turnover rates from 70% per year to 120%. They outperformed the market on average by 2.35% before switching and were outperformed by the market by 3.5% after switching. Amateur traders clearly underperformed the market and the most active traders experienced the worst performance. While commissions certainly hurt these traders, the data suggested that bid-ask spreads hurt them even more. However, their trading activity would have reduced their wealth levels even if trading costs had been zero. Interestingly, investors who traded the least actually beat market indices.

Overconfidence extends beyond the individual investor. For example, Hilary and Menzly (2006) argue that professional stock analysts who have outperformed their peers in the recent past tend to become overconfident and underperform their peers in subsequent periods. In another line of inquiry, Lichtenstein, Fischhoff, and Phillips (1982) argue that people tend to be overconfident in their own judgments. Griffin and Tversky (1992) argue that experts tend to be more prone to overconfidence than novices. Taylor and Brown (1988) argue that overconfidence leads "to higher motivation, greater persistence, more effective performance and ultimately more success." Perhaps, according to Kyle and Wang (1997), more aggressive trading behavior of overconfident professional traders enables them to generate higher profits than their more rational competitors.

Overconfidence and Gender

The relationship between overconfidence and performance is rather complicated. First, experimental evidence presented by Deaves, Lüders, and Luo (2009) suggest that overconfidence, as measured by degree of miscalibration (perhaps indicated by performance on the trivia quiz above) leads to increased trading activity. However, they did not find a correlation between gender and degree of miscalibration. Nonetheless, other studies such as Barber and Odean (2001a,b) have suggested that men are more prone to overconfidence than women, particularly in male-dominated realms such as finance. Dorn and Huberman (2005) found that investors who believe that they are better than average trade more

frequently. Using account data from over 35,000 accounts from a discount brokerage firm, Barber and Odean (2001a,b) find that men trade 45% more frequently than women, reducing their returns relative to market indices by 2.65% compared to 1.72% for women. Differences between men and women in the trading realm are so striking that one might ask whether people trade for entertainment in addition to wealth creation.

Evolutionary biologists have argued that males of many species tend to take more risks than their female counterparts (e.g., Gat, 2000). Gat suggests that males take increased risks to enhance their status in order to create more opportunities to reproduce, knowing that prospective mates prefer higher-status males.

If individual traders, particularly aggressive individual traders, lose money relative to the market, who makes money? Consider a study by Barber, Lee, and Liu (2009) covering the entire Taiwanese stock market from 1995 to 1999. They documented that individual investor trading results in consistently large losses averaging approximately 3.8%. These individual investor losses amount to 2.2% of Taiwan's GDP, almost as much as the country's total private expenditures on clothing and footwear (Barber et al., 2008). These losses are attributable to aggressive trading behavior. On the other hand, Barber et al. found that institutional investors outperform the market by 1.5% (after commissions and taxes, but before other costs passed on to their retail investors). Both aggressive and passive trades of institutions are profitable. Barber et al. (2008) conclude that "the trading and market-timing losses of individual investors represent gains for institutional investors."

Entertainment and Sensation-Seeking

Aggressive trading behavior does seem related to overconfidence, but there is also good reason to think that it can be related to sensation or thrill seeking, just as gambling might be. For example, Barber et al. (2008) argued that "the introduction of a national lottery in Taiwan reduced trading on the Taiwan Stock Exchange by about one fourth." Dorn, Dorn, and Sengmueller (2015) made a similar argument based on evidence arising from California, multistate U.S., and German lotteries, showing that individual traders reduced trading participation during weeks when lotteries such as Powerball and Mega Millions advertised increased jackpots. Based on survey data and actual trading records, Dorn and Sengmueller (2009) found, based on a survey of 1000 German brokerage clients, that respondents who professed to enjoy trading or gambling turn over their portfolios 1.5 times as often as respondents who profess no such enjoyment.

In a paper linking thrill-seeking behavior with aggressive trading, Grinblatt and Keloharju (2009) demonstrate that aggressive trading is directly related to the number of speeding tickets that traders receive. In earlier research, Zuckerman (1994) and Jonah, Thiessen, and Au-Yeung (2001) suggested that driving behavior is an excellent observable behavior for assessing thrill seeking. Furthermore, the finding that men trade more aggressively may be rooted in the same quality that leads them to more frequently violate speed limits. Men may simply enjoy the thrill of trading and the thrill of speeding more than women appear to.

Sensation-seeking motivations and behavior appear to extend beyond individual traders into institutional investor realms, including hedge funds. Brown et al. (2018) compiled

data that enabled them to match 1774 personal vehicles, of which 163 vehicles were sports cars, to 1144 U.S.-based hedge fund managers. Their tests indicated that sports car–owning hedge fund managers took on greater investment risk than other hedge fund managers. However, these apparently sensation-seeking managers failed to provide commensurate returns, leading to reduced Sharpe ratios and reduced fund alphas (see Chapter 10 of this book). Sensation-seeking managers traded more frequently and purchased higher-risk shares. Furthermore, sports car–owning managers were more likely to terminate their funds and report regulatory, civil, and criminal violations on their Forms ADV (Uniform Application for Investment Adviser Registration and Report). Brown et al. suggest that while some sensation-seeking investors may actually have sought such sensation-seeking managers, their tests were unable to link sports car purchases with ability to attract higher levels of capital.

Professional Traders and Hormones

It is well-known that testosterone and cortisol, hormones more abundant in the male body than in the female, have clear cognitive and behavioral effects. For example, testosterone is more prevalent in the bodies of winning male athletes than in losing athletes, and cortisol is known to increase in situations characterized by uncontrollability, novelty, and uncertainty. But, what is the relationship between these hormones and trading behavior? Coates and Herbert (2008) sampled, under actual working conditions, endogenous steroids from a group of male London traders in the City of London. They found that a trader's morning testosterone level predicts his day's trading profitability. More specifically, they found that on mornings when testosterone levels were high, 14 of the 17 traders in their study realized higher trading profits. They also found that a trader's cortisol rises with both the variance of his trading results and the volatility of the market. Thus, Coates and Herbert found that that higher testosterone levels seem to contribute to trading returns while cortisol levels increased as risk and risk-taking increase.

Coates and Herbert expressed concern that chronically elevated testosterone levels could have negative effects on returns, because testosterone has also been found to lead to impulsivity and sensation seeking along with harmful risk taking. For example, in one study, testosterone administered to players of a gambling game led to their accepting irrational risk/reward tradeoffs.

It is clear that women and men show differences in their investment decision-making. Neyse et al. (2016), in their experimental studies, and studies that we have already discussed find that male subjects are more overoptimistic than females in their prediction abilities. Women are more risk averse and men trade more aggressively. In large-sample studies of brokerage firm clienteles (amateur traders), women outperform men. Here, we examine more closely certain physical differences between men and women along with their trading behavior in more professional trading settings. Hormonal influences in particular might have important effects on market behavior.

In their study of experimental computerized double-auction asset markets involving 142 "traders," Cueva et al. (2015) show that both cortisol, which elevates in response to physical or psychological stress (risk), and testosterone levels, which are predictive of

confidence and success in a variety of competitive circumstances, shifted the asset selection of young males towards riskier instruments. They argued that the presence of these hormones at elevated levels could play destabilizing roles in financial markets. In their experiments, relevant hormones were either externally induced (elevated by pills or injection) or natural levels were simply measured in saliva. Nevertheless, higher levels of these hormones were clearly related to risk-taking in males.

While androgenic hormone levels are relevant to our competitive performance, even exposure in utero to these hormones seems to be relevant. Consider, for example, the ratio 2D/4D of index finger (2D) length to ring finger (4D) length, which numerous studies have suggested to be inversely correlated with prenatal testosterone exposure (e.g., Williams et al., 2000).[12] The index to ring finger length ratio is inversely correlated to both improved confidence and improved performance in a variety of competitive activities ranging from sports to trading (e.g., Coates, Gurnell, & Rustichini, 2009; Manning, 2002). That is, first, males tend to have lower index to ring finger (2D/4D) ratios than females, and lower 2D to 4D ratios are associated with both higher testosterone exposure in utero and later levels of confidence and performance in competitive situations. In addition to realizing superior trading performance, males with lower digit ratios remain in the market longer.

Coates, Gurnell, and Sarnyai (2010) suggest that, while markets might, in general reflect all available information (be efficient), traders' interpretation of this information might not be stable, in part, due to fluctuating hormone levels. High levels of testosterone tend to increase trading confidence while increasing levels of cortisol may decrease risk-taking. If such hormone levels vary systematically across traders, market stability might be affected. Diversity, particularly gender diversity on trading floors has the potential to mitigate the influence of systematic hormonal effects on market behavior.

11.5 NEUROFINANCE: GETTING INTO THE INVESTOR'S HEAD

While neoclassical economics has given us excellent tools to understand market behavior, behavioral finance has enabled us to better understand how financial markets might behave when its participants are irrational. Behavioral finance uses methods and results from cognitive psychology to better understand financial decision making. *Neurofinance*, in its infancy stages, is concerned with understanding the neurological processes in the investor's brain as she or he makes financial decisions. We already discussed studies concerning potential physiological causes of hyperbolic discounting and concerning the effects of testosterone on trading. In another well-publicized paper, Shiv, Lowenstein, Bechara, Damasio, and Damasio (2005) studied the relative abilities of brain-damaged study participants to make gambling decisions. This study gathered 19 subjects who had incurred damage (stable focal lesions) to parts of their brains impairing their abilities to process emotions. The subjects were asked to participate in a series of gambles along with two control groups, one that had experienced

[12]Measure ring and index fingers from the tip to the skin crease at the base to obtain their lengths. Then, divide the index finger length by the ring finger length to obtain the index ratio. Male ratios average 2D/4D = 0.947 ($\sigma = 0.029$) and females 0.965 ($\sigma = 0.026$) (Bailey & Hurd, 2005).

no brain damage and a second group that had experienced some other type of brain damage. Each study participant was asked to participate in a sequential series of 20 gambles, betting $1 against a 50/50 chance at either 0 or $2.50. The expected value of each gamble was $1.25, $0.25 higher than its cost. The subjects experiencing damage to their emotional circuitry bet more consistently than their "normal" counterparts and earned more money. The performance differences were more pronounced after nonimpaired subjects experienced losses, making them even more reluctant to take advantage of expected wealth-increasing gambles. The performance of the emotionally damaged group compared favorably to the control group of participants who had experienced no brain damage and to the second control group who had experienced unrelated types of brain damage.

A contrasting study by Naqvi, Shiv, and Bechara (2006) found that subjects with similar brain damage (in the ventromedial prefrontal cortex) impairing their abilities to experience emotion seem unable to learn from mistakes in everyday life decisions, leading to repeated impairment of their well-being. Similarly, when faced with repeated losses in "rigged" gambling scenarios, subjects with impaired ability to experience emotions seemed unable to learn from negative experiences. Perhaps, in sum, this and the previous studies suggest that emotions are useful in reacting to negative experiences but can lead to irrational overreactions.

Other neurofinance studies have focused on brain imaging tools, including Electroencephalography (EEG), positron emission tomography (PET) scans and, most importantly, functional magnetic resonance imaging (fMRI), which is used to localize the presence of oxygen-rich blood. These imaging techniques allow researchers to examine the activity in the subject's brain as the subject experiences various stimuli, such as facing risk. For example, Knutson and Peterson (2005) discuss the application of fMRI technology to map expected utility while Kuhnen and Knutson (2005) focus directly on financial risk taking. In one study of professional foreign exchange traders, Lo and Repin (2002) used fMRI to find that more experienced traders experienced significantly less emotional reaction to dramatic market changes than did their less experienced counterparts. Essentially, they "wired" traders to examine their physiological reactions to changes in market conditions.[13] Other studies have made use of intra- and extra-cellular recording of electrical activity, which involves inserting electrodes into the cell bodies of neurons.

Neurofinance seems to be a very promising approach to understanding the way investors make decisions. Most financial research is based on the methodologies from classical economics where economists model how investors ought to behave in order to maximize their wealth, and then examine statistical and case data to determine whether they do behave in the manners that the models predict. While this study methodology has proven to be very successful in most respects, it leaves many important questions unanswered. Behavioral finance research often examines how investors make decisions in experimental or laboratory settings. Neurofinance pursues the same questions by examining the physical biological reactions of subjects to stimuli.

[13]Zak (2004) provides a very readable introduction to neuroeconomics methodology and literature and Peterson (2007) provides an introduction more specific to finance literature.

11.6 THE CONSENSUS OPINION: STUPID INVESTORS, SMART MARKETS?

Is it possible for a market comprised of irrational investors to actually, in sum, behave rationally? Is it likely that rational or educated investors play a greater role in the price-setting process than less rational or poorly educated investors? Could poorly educated or irrational investors be driven out of the price-setting process as their wealthier investors gain more power as they accumulate more wealth at the expense of their irrational counterparts? Consider a hypothetical market where professional analysts and competing investors are attempting to secure and employ all information that would enable them to evaluate stocks more accurately. However, none of the analysts have perfect information, and cannot know with certainty what stock values are. Further assume that each analyst may have some information (or method for analyzing this information) not available to other analysts. However, each analyst may be lacking some information or technique known to his competitors. Thus, information sets available to different analysts are not perfectly correlated. Given a reasonably large number of analysts, one might expect their errors to offset or cancel to some extent and that their "average" or consensus projections to outperform any given analyst's forecasts. In this scenario, the market prices generated by the large number of analysts and investors may be closer to the true security prices estimated by any individual analyst or investor. It would be very difficult for the individual investor to beat the market in this scenario.

Surowiecki (2004) described the popular TV show *Who Wants To Be a Millionaire?* to demonstrate the "wisdom of crowds" relative to individual decision makers. In this show, a contestant was asked multiple-choice questions, which, if answered correctly, could result in winnings of as much as $1 million. The contestant had the option ("lifeline") of seeking each of three types of assistance should he require it. The contestant could request to have two of three incorrect answers eliminated from the answer set, call a friend or relative to ask for help, or poll the studio audience to vote on the correct answer. Eliminating incorrect answers should produce correct answers at least 50% of the time. Phone calls to friends or relatives produced the correct answer almost 65% of the time. However, the studio polls produced the correct answers 91% of the time, suggesting that crowd wisdom did seem superior to individual opinions, even the potentially expert opinions offered by the phone calls.

Surowiecki proposed conditions under which collective wisdom works well, including diversity, independence, and aggregation. For example, investors have diverse decision rules and may well make investment decisions independently of one another. They absorb information from their environments, combine it with their own interactions with their environments, and then formulate decision rules. Their effective decision rules survive, adapt, and are maintained. Securities markets provide the mechanism for aggregating the diverse rules and opinions of investors. De Figueiredo and de Figueiredo (2002) demonstrated with experimental evidence that while individual managers are able to effectively resolve less complicated problems, markets provide better settings for the resolution of more complex issues.

Numerous experiments have demonstrated that averages of classroom estimates of temperatures are more accurate than individual student estimates. Similarly, average

estimates provided by surveys produce better estimates of numbers of jellybeans in jars than individual estimates. Such experiments do suggest that the crowd may compile information to produce better estimates than individuals acting alone.

The Football Pool

Sports forecasting and betting provide excellent opportunities for testing market efficiency in that true outcomes are revealed after games are played. Consider data in Table 11.4 based on a paper by Beaver (1981) to describe how football reporters might find it difficult to outperform consensus forecasts made by a group of their colleagues.[14] The *Chicago Daily News* recorded the college football predictions of its sports staff for the last weekend of November during the 1966 to 1968 seasons. While the results of this "study" are not intended to be suitable for scientific purposes, they do illustrate the difficulty faced by forecasters attempting to outperform their peers.

While results reported in Table 11.4 do suggest that a consensus generates better football predictions than a single analyst, to what extent might this result hold in the stock market? An individual analyst might be regarded as able to generate superior forecasts if his returns consistently outperform those of the market on a risk-adjusted basis. While it is certain that at least a small number of investors consistently outperform the market, it is likely that a much larger number simply claim to do so. Furthermore, in a market with millions of investors, many would consistently outperform the market even if all selected their investments randomly. Perhaps, one should be somewhat skeptical when reviewing reports of superior investment performance.

TABLE 11.4 The Football Pool

	1966	**1967**	**1968**
Total forecasters (including consensus)	15	15	16
Total forecasts made per forecaster	180	220	219
Rank of consensus forecast	1-tie	2	2
Median rank of forecasters	8	8	8.5
Rank of best forecasters			
J. Carmichael (1966)	1-tie	8	16
D. Nightingale (1966)	1-tie	11	5
A. Biondo (1967)	7	1	6
H. Duck (1968)	8	10	1

Sources: Beaver (1981), figure 6.1, 162; Foster (1986), table 9.1, 325.find.

[14]This example was described in Foster (1986).

Analyst Estimates

Because earnings forecasts are such an integral part of the typical analyst's stock valuation process, a number of studies have been performed to determine analysts' abilities to forecast earnings per share (EPS). Studies have indicated that consensus forecasts for EPS are superior to those of a randomly selected analyst (e.g., Fried and Givoly, 1982). By combining a large number of forecasts, individual analyst idiosyncratic errors will tend to offset one another. Several firms make consensus forecasts available to the public, including Lynch, Jones & Ryan's Institutional Brokers Estimate System (IBES), First Call (a subsidiary of Thomson Corporation), and Zacks Investment Research, Inc. Furthermore, as one might expect, studies suggest that EPS forecasts become more accurate as announcement dates approach (e.g., Brown & Chen, 1990). However, a number of studies have demonstrated that analysts' EPS forecasts tend to be overly optimistic (e.g., Barber, Lehavy, McNichols, & Trueman, 2006), perhaps by as much as 20%. A number of hypotheses have been offered to explain this phenomenon, including interests of the underwriting divisions of investment banks. Nonetheless, there is evidence that investors account for this bias and are able to sift through a huge volume of other information (and misinformation) to price securities. The extent to which investors and the market are able to price securities appropriately represents the extent to which the market is efficient.

Rational Investors and Price Setting

Consider again our earlier example with a minimum of two rational traders and many other irrational investors (as in Kluger & Wyatt, 2004). The two rational traders competing against others can force prices up to rational levels as they compete to buy. They will force prices down to rational levels as they compete to sell. As irrational buyers buy at high prices, rational sellers will sell, competing to keep prices down to rational levels. As irrational buyers sell at low prices, rational sellers will compete to buy, competing to keep prices up to rational levels. A somewhat similar finding by Schwartz (2008) that firms forced to compete for the business of both sophisticated and naïve customers are less likely to exploit less sophisticated customers than firms that don't compete. In addition, List (2003) found that, in markets for sports memorabilia, the "endowment effect" (such as those experienced by "Joe and his opera tickets" or the roulette wheel experiments) disappears as dealers' levels experience increase.

Observations of the behavior of individuals and a variety of experiments have indicated that apparently irrational decision making frequently governs actions of consumers and investors, particularly in cases that involve potential losses. Nonetheless, might one expect that rational investors will outperform irrational investors in the market? Might rational investors tend to better preserve their capital, and even that of other investors (e.g., as professional money managers)? Perhaps, the market will tend to drive out (or bankrupt) irrational investors or at least, in time, reduce their impact on the market? Perhaps the rational investors, always driving to exploit an opportunity at the expense of irrational investors, will force market prices to properly reflect information whenever a deviation occurs? That is, might rational investors through their own efforts to secure profits, dominate the market and force prices to fully and rationally reflect available information?

Furthermore, people observing the actions of others may be all too eager to note apparent irrational behavior. One is usually not able to observe the actual rationale or basis for the actions of others, and hence, may not be in a strong position to determine whether the resulting behavior truly is rational.

Herds and Swarms

Stock markets are key mechanisms for planning and allocating economic resources for society. However, unlike governments, military organizations, and corporations, they function without formal leadership or hierarchies. No one is in charge of price setting in stock markets and determining who owns what securities in the investment arena. In fact, some observers have compared stock markets to swarms of bees and ant colonies. Miller (2007, 130) wrote that "Ants aren't clever little engineers, architects, or warriors after all—at least not as individuals." He quotes the biologist Deborah Gordan, "If you watch an ant try to accomplish something, you'll be impressed by how inept it is. Ants aren't smart. Ant colonies are." Ants collectively decide how, when, and where to forage for food. As Miller writes, they are "simple creatures following simple rules, each one acting on local information. No ant tells any other ant what to do.... No leadership is required. Even complex behavior may be coordinated by relatively simple interactions." Each ant has a tiny sliver of information that is communicated in a very rudimentary fashion to other ants, but no ant comes close to understanding the "big picture" and no ant can direct the activities of the colony as a whole. Nevertheless, a huge ant colony with hundreds of thousands of ants can thrive, feed itself, reproduce, take care of its young, fight, and even enslave other species.

The stock market may function quite similarly. Individual traders, each with a tiny subset of relevant information communicates bids and offers. Most traders do not reveal their rationale for their quotations, only their quotes. Through this process, the market collectively sets prices and allocates productive resources throughout the economy. All of this is accomplished without formal leadership or without anyone really understanding exactly why stock prices behave as they do. Ivković and Weisbenner (2007) find evidence of herding in an examination of 35,000 brokerage accounts detailing investor zip codes. They found that investors were substantially more likely to invest in securities if their neighbors had already done so. The likelihood of investing in a given stock dropped substantially as the distance to other investors who had already done so increased. To deal with the potential problem of investors simply buying stocks or in industries important to the local economy, Ivković and Weisbenner used a "social interaction barometer" developed by Putnam (2000) and found that the herding tendency was greatest in states with higher levels of social interaction. However, markets (and ant colonies as well) may be capable of committing enormous collective blunders, which might be termed bubbles and crashes.

An *information cascade* might be characterized as a sequential decision process in which each decision maker bases his decision on those made by previous decision makers and then follow what their predecessors did in the decision-making queue did rather than use his own information. This information cascading may form the basis for

herd behavior, where decision makers pursue the same action without collaborative planning. In an information cascade, decision makers earlier in the queue have information relevant to subsequent decision makers, so this herding might be rational. The sequential nature of this decision making and its information flow is what characterizes information cascading. But, herd behavior need not be the result of sequential decision making. Members of a herd merely need to exhibit the same behavior without collaborative planning. In fact, herding behavior may seem consistent with collective irrationality. Herding behavior has been blamed for stock market bubbles and crashes as well as bank runs.

Additional Reading

Nofsinger's 2014 book on the psychology of investing is interesting reading and very accessible to a beginner. Kahneman and Tversky (1982) contains a collection of their classic essays (other authors contributed as well) that serve as foundation classic reading for this field. Lewis (2016), in a highly readable book, chronicles the long friendship and collaboration between Kahneman and Tversky and provides numerous detailed illustrations of their findings. Shiv et al.'s 2005 scientific article on the role of emotion in investment decision making drew substantial coverage in the popular and financial press, and Zak (2004) provides a review of academic literature in the neurofinance field. Hirshleifer and Teoh (2003) provide an excellent review of the herding literature, including discussions of herding in stock markets, bank runs, creditor runs, analyst herding, and herding in accounting reports. More generally, Surowiecki's 2004 book is highly readable, very entertaining, and offers striking examples concerning crowd wisdom, both in and outside the finance field. Statman (1997) and Ritter (2003) provide good overviews of the academic literature in behavioral finance, while Shefrin's 2002 book is an excellent and very readable introduction to the field. Miller's 2007 article is a highly readable and thoroughly entertaining discussion of swarm behavior in the insect kingdom, and describes parallels between swarm behavior in the insect kingdom and the stock market.

References

Bailey, A. A., & Hurd, P. L. (2005). Finger length ratio (2D:4D) correlates with physical aggression in men but not in women. *Biological Psychology*, *68*(3), 215–222.

Barber, B., & Odean, T. (2000). Trading is hazardous to your wealth: The common stock investment performance of individual investors. *Journal of Finance*, *55*, 773–806.

Barber, B., & Odean, T. (2001a). Boys will be boys: Gender, overconfidence, and common stock investment. *Quarterly Journal of Economics*, *116*(1), 261–292.

Barber, B., & Odean, T. (2001b). On-line investors: Do the slow die first? *Review of Financial Studies*, *15*(2), 455–487.

Barber, B., Lee, Y. T., & Liu, Y. J. (2009). Just how much do investors lose from trade? *Review of Financial Studies*, *22*, 609–632.

Barber, B. M., Lee, Y. T., Liu, Y. J., & Odean, T. (2008). Just how much do individual investors lose by trading? *Review of Financial Studies*, *22*, 609–632.

Barber, B. M., Lee, Y. T., Liu, Y. J., Odean, T., & Zhang, K. (2020). Learning, fast or slow. *The Review of Asset Pricing Studies*, *10*(1), 61–93.

Barber, B., Lehavy, R., McNichols, M., & Trueman, B. (2006). Buys, holds and sells: The distribution of investment banks' stock ratings and the implications for the profitability of analysts' recommendations. *Journal of Accounting and Economics*, *41*, 87–117.

Baron, J. (2000). *Thinking and deciding* (3rd ed.). Cambridge: Cambridge University Press.

Barraclough, K., & Whaley, R. E. (2012). Early exercise of put options on stocks. *Journal of Finance*, *67*(4), 1423–1456.

Beattie, V., & Jones, M. (2000). Impression management: The case of inter-country financial graphs. *Journal of International Accounting, Auditing & Taxation*, *9*, 159–183.
Beaver, W. H. (1981). *Financial reporting: An accounting revolution*. Englewood Cliffs, NJ: Prentice-Hall.
Benartzi, S., & Thaler, R. (1995). Myopic loss-aversion and the equity premium puzzle. *Quarterly Journal of Economics*, *110*, 73–92.
Black, F. (1986). Noise. *Journal of Finance*, *41*, 529–544.
Brown, J. R., Kapteyn, A., Erzo, F. P., Luttmer, E. F. P., & Mitchell, O. S. (2017). Cognitive constraints on valuing annuities. *Journal of the European Economic Association*, *15*(2), 429–462.
Brown, L. D., & Chen, D. M. (1990). Composite analysts earnings forecasts: The next generation. *Journal of Business Forecasting*, *9*, 11–15.
Brown, S., Lu, Y., Ray, S., & Teo, M. (2018). Sensation seeking and hedge funds. *Journal of Finance*, *73*(6), 2871–2914.
Coates, J. M., & Herbert, J. (2008). Endogenous steroids and financial risk taking on a London trading floor. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 6167–6172.
Coates, J. M., Gurnell, M., & Rustichini, A. (2009). Second-to-fourth digit ratio predicts success among high-frequency financial traders. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 623.
Coates, J. M., Gurnell, M., & Sarnyai, Z. (2010). From molecule to market: Steroid hormones and financial risk-taking. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 331.
Coval, J., & Moskowitz, T. (1999). Home bias at home: Local equity preference in domestic portfolios. *Journal of Finance*, *54*, 2045–2074.
Coval, J., & Moskowitz, T. (2001). The geography of investment: Informed trading and asset prices. *Journal of Political Economy*, *109*, 811–841.
Coval, J. D., Hirshleifer, D., & Shumway, T. (2005). *Can individual investors beat the market*? Unpublished working paper No. 04-025. School of Finance, Harvard University.
Cueva, C., Roberts, R. E., Spencer, T., Rani, N., Tempest, M., Tobler, P. N., et al. (2015). Cortisol and testosterone increase financial risk taking and may destabilize financial markets. *Scientific Reports*, *5*, 11206.
De Figueiredo, J., & de Figueiredo, R. J., Jr. (2002). Managerial decision-making in non-market environments: A survey experiment. *Advances in Strategic Marketing*, *67–96*.
Deaves, R., Lüders, E., & Luo, G. Y. (2009). An experimental test of the impact of overconfidence and gender on trading activity. *Review of Finance*, *13*(3), 555–575.
DeBondt, W. F. M., & Thaler, R. (1985). Does the market overreact? *Journal of Finance*, *40*, 793–805.
Diacon, S., & Hasseldine, J. (2007). Framing effects and risk perception: The effect of prior performance presentation format on investment fund choice. *Journal of Economic Psychology*, *28*, 31–52.
Dichev, I. D., & Janes, T. D. (2001). *Lunar cycle effects in stock returns*. Working Paper Series. Available at <http://ssrn.com/abstract = 281665> or <http://dx.doi.org/10.2139/ssrn.281665>.
Dimson, E., Marsh, P., & Staunton, M. (2002). *Triumph of the optimists: 101 years of global investment*. Princeton, NJ: Princeton University Press.
Dorn, A. J., Dorn, D., & Sengmueller, P. (2015). Trading as gambling. *Management Science*, *61*(10), 2281–2547.
Dorn, D., & Huberman, G. (2005). Talk and action: What individual investors say and what they do. *Review of Finance*, *9*, 437–481.
Dorn, D., & Sengmueller, P. (2009). Trading as entertainment? *Management Science*, *55*(4), 591–603.
Edmans, A., García, D., & Norli, Ø. (2007). Sports sentiment and stock returns. *Journal of Finance*, *62*(4), 1967–1998.
Foster, G. (1986). *Financial statement analysis*. Englewood Cliffs, NJ: Prentice-Hall.
Frazzini, A. (2006). The disposition effect and underreaction to news. *The Journal of Finance*, *61*(4), 2017–2046.
Fried, D., & Givoly, D. (1982). Financial analysts' forecasts of earnings: A better surrogate for market expectations. *Journal of Accounting and Economics*, *4*(2), 85–107.
Friedman, D. (1998). Monty Hall's three doors: Construction and deconstruction of a choice anomaly. *American Economic Review*, *88*, 933–946.
Gat, A. (2000). The human motivational complex: Evolutionary theory and the causes of hunter-gatherer fighting. *Anthropological Quarterly*, *73*, 20–34.
Genesove, D., & Mayer, C. (2001). Loss aversion and seller behavior: Evidence from the housing market. *Quarterly Journal of Economics*, *116*(4), 1233–1260.

Gneezy, U., Kapteyn, A., & Potters, J. (2003). Evaluation periods and asset prices in a market experiment. *Journal of Finance*, *58*, 821–837.
Goetzmann, W. N., & Kumar, A. (2008). Equity portfolio diversification. *Review of Finance*, *12*, 433–463.
Goetzmann, W. N., & Zhu, N. (2005). Rain or shine: Where is the weather effect? *European Financial Management*, *11*, 559–578.
Griffin, D., & Tversky, A. (1992). The weighing of evidence and the determinants of confidence. *Cognitive Psychology*, *24*, 411–435.
Grinblatt, M., & Han, B. (2005). Prospect theory, mental accounting, and momentum. *Journal of Financial Economics*, *78*, 311–339.
Grinblatt, M., & Keloharju, M. (2009). Sensation seeking, overconfidence, and trading activity. *Journal of Finance*, *64*(2), 549–578.
Grinblatt, M., Keloharju, M., & Linnainmaa, J. T. (2011). IQ and stock market participation. *Journal of Finance*, *66*(6), 2121–2164.
Hilary, G., & Menzly, L. (2006). Does past success lead analysts to become overconfident? *Management Science*, *52*, 489–500.
Hirshleifer, D., & Shumway, T. (2003). Good day sunshine: Stock returns and the weather. *Journal of Finance*, *58*, 1009–1032.
Hirshleifer, D., & Teoh, T. H. (2003). Herd behavior and cascading in capital markets: A review and synthesis. *European Financial Management*, *9*(1), 25–66.
Ito, T. (1990). Foreign exchange expectations: Micro survey data. *American Economic Review*, *80*, 434–449.
Ivković, Z., & Weisbenner, S. (2005). Local does as local is: Information content of the geography of individual investors' common stock investments. *Journal of Finance*, *60*, 267–306.
Ivković, Z., & Weisbenner, S. (2007). Information diffusion effects in individual investors' common stock purchases: Covet thy neighbors' investment choices. *Review of Financial Studies*, *20*(4), 1327–1357.
Ivković, Z., Poterba, J., & Weisbenner, S. (2005). Tax-motivated trading by individual investors. *American Economic Review*, *95*(5), 1605–1630.
Ivković, Z., Sialm, C., & Weisbenner, S. (2008). Portfolio concentration and the performance of individual investors. *Journal of Financial and Quantitative Analysis*, *43*, 613–655.
Jonah, B. A., Thiessen, R., & Au-Yeung, E. (2001). Sensation seeking, risky driving and behavioral adaptation. *Accident Analysis & Prevention*, *33*(5), 679–684.
Kadapakkam, P. R., & Zhang, H. (2014). Investor ignorance in markets for worthless stocks. *Journal of Financial Markets*, *19*, 197–218.
Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263–291.
Kahneman, D., & Tversky, A. (1981). The framing of decisions and the psychology of choice. *Science*, *211*(30), 453–457.
Kahneman, D., & Tversky, A. (1982). Intuitive prediction: Biases and corrective procedures. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 414–421). Cambridge: Cambridge University Press.
Kamstra, M. J., Kramer, L. A., & Levi, M. D. (2003). Winter blues: A SAD stock market cycle. *American Economic Review*, *93*(1), 324–343.
Kim, J. H. (2016). *Stock returns and investors' mood: Good day sunshine or spurious correlation?* Unpublished Working paper, La Trobe University.
Kluger, B. D., & Wyatt, S. B. (2004). Are judgment errors reflected in market prices and allocations? Experimental evidence based on the Monty Hall problem. *Journal of Finance*, *59*(3), 969–997.
Knutson, B., & Peterson, R. (2005). Neurally reconstructing expected utility. *Games and Economic Behavior*, *52*, 305–315.
Kuhnen, C., & Knutson, B. (2005). The neural basis of financial risk taking. *Neuron*, *47*, 763–770.
Kyle, A. S., & Wang, F. A. (1997). Speculation duopoly with agreement to disagree: Can overconfidence survive the market test? *Journal of Finance*, *52*(5), 2073–2089.
Laibson, D. (1997). Golden eggs and hyperbolic discounting. *Quarterly Journal of Economics*, *112*, 443–477.
Lewis, M. (2016). *The undoing project: A friendship that changed our minds*. New York: W. W. Norton.

Liao, L., Chou, R. Y., & Chiu, B. (2013). Anchoring effect on foreign institutional investors' momentum trading behavior: Evidence from Taiwan stock market. *North American Journal of Economics and Finance*, *26*, 72–91.

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 306–334). Cambridge: Cambridge University Press.

Limpaphayom, P., Locke, P. R., & Sarajoti, P. (2007). Gone with the wind: Chicago's weather and futures trading. *Review of Futures Markets*, *16*, 1.

List, J. A. (2003). Does market experience eliminate market anomalies? *Quarterly Journal of Economics*, *118*(1), 41–71.

Lo, A. W., & Repin, D. V. (2002). The psychophysiology of real-time financial risk processing. *Journal of Cognitive Neuroscience*, *14*(3), 323–339.

Manning, J. T. (2002). *Digit ratio: A pointer to fertility, behavior, and health*. New Brunswick: Rutgers University Press.

Massa, M., & Simonov, A. (2006). Hedging, familiarity, and portfolio choice. *Review of Financial Studies*, *19*(2), 633–685.

Miller, P. (2007). Swarm theory. *National Geographic*, 126–147.

Naqvi, N., Shiv, B., & Bechara, A. (2006). The role of emotion in decision-making: A neuroscience perspective. *Current Directions in Psychological Science*, *15*(4), 260–264.

Neyse, L., Bosworth, S., Ring, P., & Schmidt, U. (2016). Overconfidence, incentives and digit ratio. *Scientific Reports 6, Article number*, *23294*.

Nofsinger, J. (2014). *The psychology of investing* (5th ed.). New York: Routledge.

O'Donoghue, T., & Rabin, M. (2006). Optimal sin taxes. *Journal of Public Economics*, *90*(10–11), 1825–1849.

Odean, T. (1998a). Are investors reluctant to realize their losses? *Journal of Finance*, *53*(5), 1775–1798.

Odean, T. (1998b). Volume, volatility, price and profit when all traders are above average. *Journal of Finance*, *53*(6), 1887–1934.

Peterson, R. (2007). Affect and financial decision-making: How neuroscience can inform market participants. *Journal of Behavioral Finance*, *8*(2), 1–9.

Pool, V. K., Stoll, H. R., & Whaley, R. E. (2008). Failure to exercise call options: An anomaly and a trading game. *Journal of Financial Markets*, *11*, 1–35.

Poteshman, A. M., & Serbin, V. (2003). Clearly irrational financial market behavior: Evidence from the early exercise of exchange traded stock options. *Journal of Finance*, *58*, 37–70.

Putnam, R. D. (2000). *Bowling alone*. New York: Simon and Schuster.

Rashes, M. S. (2001). Massively confused investors making conspicuously ignorant choices (MCI-MCIC). *Journal of Finance*, *56*(5), 1911–1928.

Ritter, J. (2003). Behavioral finance. *Pacific Basin Finance Journal*, *11*(4), 429–437.

Russo, J. E., & Schoemaker, P. J. H. (1992). Managing overconfidence. *MIT Sloan Management Review*, *33*(2), 7–17.

Schwartz, A. (2008). How much irrationality does the market permit? *Journal of Legal Studies*, *37*, 131.

Shefrin, H. (2002). *Beyond greed and fear: Understanding behavioral finance and the psychology of investing*. Oxford: Oxford University Press.

Shefrin, H., & Statman, M. (1985). The disposition to sell winners too early and ride losers too long: Theory and evidence. *Journal of Finance*, *40*, 777–790.

Shiller, R. J. (1981). Do stock prices move too much to be justified by subsequent changes in dividends? *American Economic Review*, *71*, 421–436.

Shiller, R. J., Kon-Ya, F., & Tsutsui, Y. (1996). Why did the Nikkei crash? Expanding the scope of expectations data collection. *Review of Economics and Statistics*, *78*(1), 156–164.

Shiv, B., Lowenstein, G., Bechara, A., Damasio, H., & Damasio, A. (2005). Investment behavior and the negative side of emotion. *Psychological Science*, *16*(6), 435–439.

Sorescu, S., & Subramanyam, A. (2006). The cross section of analyst recommendations. *Journal of Financial and Quantitative Analysis*, *41*(1), 139–168.

Statman, M. (1997). Behavioral finance. *Contemporary Finance Digest*, *1*(1), 5–21.

Strotz, R. H. (1956). Myopia and inconsistency in dynamic utility maximization. *Review of Economic Studies*, *23*, 165–180.

Surowiecki, J. (2004). *The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economics, societies, and nations*. New York: Doubleday.
Svenson, O. (1981). Are we all less risky and more skilled than our fellow drivers? *Acta Psychologica*, *47*, 143–148.
Taylor, S., & Brown, J. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin*, *103*, 193–210.
Thaler, R. H., Tversky, A., Kahneman, D., & Schwartz, A. (1997). The effect of myopia and loss aversion on risk-taking: An experimental test. *The Quarterly Journal of Economics*, *112*(2), 647–661.
Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*, 453–458.
Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *Journal of Business*, *59*(4), 251–278, Part 2: The Behavioral Foundations of Economic Theory. (October 1986), S251–S278.
von Neumann, J., & Morgenstern, O. (1947). *The theory of games and economic behavior* (2nd ed.). Princeton, NJ: Princeton University Press.
Williams, T. J., Pepitone, M. E., Christensen, S. E., Cooke, B. M., Huberman, A. D., Breedlove, N. J., et al. (2000). Finger-length ratios and sexual orientation. *Nature*, *404*(6777), 455–456.
Zak, P. (2004). Neuroeconomics. *Philosophical Transactions of the Royal Society B*, *359*, 1737–1748.
Zuckerman, M. (1994). *Behavioural expressions and biosocial bases of sensation-seeking*. Cambridge: Cambridge University Press.

11.7 EXERCISES

1. Suppose that an investor with \$2 in capital has a logarithmic utility of wealth function: $U = \ln(w)$. The investor has the opportunity to buy into the gamble described in the St. Petersburg paradox. Assume that the investor can borrow without interest and that the gamble payoff is 2^i where i is the number of tosses or outcomes realized before the first head is realized.
   a. What is the investor's current utility of wealth level?
   b. How much would the investor be willing to pay for the gamble described in the St. Petersburg paradox?
   c. How much would the investor be willing to pay for the gamble described in the St. Petersburg Paradox if his initial wealth level were \$1000 rather than \$2?
   d. What would be your answer to part b if the gamble payoff were to change to 2^{2i-1} where i is the number of tosses or outcomes realized before the first head is realized?

2. Three traders are at lunch and must decide on a dessert that they will share. Consider the following table, which compares trader preferences for three types of desserts, all ranked according to preferences one through three for each trader. Individual trader preferences are rational. There are three possible desserts, and thus three possible preference rankings. The traders have determined that if they vote on which dessert they will order, each of the three desserts will receive one vote, resulting in a tie. Thus, the traders decided to hold pairwise votes on combinations of two desserts. That is, they will eliminate desserts one at a time in three votes, with the winning dessert being matched against the next dessert. Thus, ice cream will be matched against cake, and then against cookies and then cookies against cake. Demonstrate that a pairwise vote (one dessert option against another) on any combination of two desserts would ultimately result in a "social" selection that will violate the principle of transitivity.

	Voter 1	*Voter* 2	*Voter* 3
Ice Cream	1	3	2
Cake	2	1	3
Cookies	3	2	1

3. A car with a replacement value of $20,000 can be insured against a total loss with an insurance policy sold for a premium of $1200. The insurance company selling the policy and the consumer purchasing the policy agree that there is a 5% probability that the car will be destroyed.
   a. What is the actuarial (fair or expected) value of the policy?
   b. If the insurance maintains a large, well-diversified portfolio of such policies, what is its expected profit from the sale of this policy?
   c. What is the expected profit (or gain or loss) to the consumer from the purchase of this policy?
   d. Under what circumstances is the sale of this policy a rational transaction for the insurance company?
   e. Under what circumstances is the purchase of this policy a rational transaction for the consumer?

4. Suppose that you are presented with the following choice:
   I. In addition to whatever you own, you have been given $1000. You are now asked to choose between:
      A: A sure gain of $500
      B: A 50% change to gain $1000 and a 50% chance to gain nothing.
   Which do you prefer: Gamble A or Gamble B? Now, suppose that you are presented with the following choice.
   II. In addition to whatever you own, you have been given $2000. You are now asked to choose between:
      A*: A sure loss of $500
      B*: A 50% chance to lose $1000 and a 50% chance to lose nothing.
   In experimental studies involving both of these sets of gambles, 84% of study participants chose A over B, however, 69% chose B* over A*. With which of the von Neumann-Morgenstern axioms would the results of this study seem inconsistent?

5. Suppose that you were to perform an experiment on subjects to determine whether they prefer to have coffee mugs or money. You plan to use a large representative sampling of individuals and endow half of the participants in your experiment with coffee mugs and then ask those to whom mugs were given what would be the lowest price at which they would sell. Those subjects who were not given mugs were asked how much they would pay for a mug.
   a. If participants are rational and consistent in their preferences, how should purchase prices differ from selling prices?
   b. Assume that the endowment effect described by Kahneman and Tversky is true for individuals. How should purchase prices differ from selling prices?
   c. What would be the relevance of this type of experiment to stock markets?

6. For each of the following three scenarios, would you rather:
   a. Receive $100 in one month or $100 in two months (most people prefer the former)?

b. Be given an excellent meal at a nice restaurant in one month or in two months (most people prefer the former)?

c. Be given a mediocre meal at a mediocre restaurant in one month and then an excellent meal at a nice restaurant in two months or an excellent meal at a nice restaurant in one month followed by mediocre meal at a mediocre restaurant in two months (most people prefer the former)?

Assume that you respond as do most people. Explain any inconsistency in your preferences.

7. Churches, schools, and other nonprofit institutions have held jelly bean counting contests where contestants compete for a prize by estimating the number of jelly beans in a large jar. There is evidence suggesting that averages of contestant estimates tend to be superior to specific individual estimates. Why might this be true?

8. In our discussion of prospect theory, we considered the following choice of gambles:

Gamble A: 0.33 probability of receiving 2500, 0.66 of receiving 2400, and 0.01 of receiving 0

Gamble B: 100% probability of receiving 2400

and

Gamble A*: 0.33 probability of receiving 2500, 0.67 of receiving 0

Gamble B*: 0.34 probability of receiving 2400 and 0.66 of receiving 0

Demonstrate that if an investor is indifferent between Gambles A and B, he must be indifferent to A* and B* in order to fulfill the strong independence axiom identified by von Neumann and Morgenstern.

9. Suppose that the United States is preparing for the outbreak of an Asian viral disease, which, if a vaccine is not developed, is expected to kill 600 people. Two alternative vaccination programs have been proposed. The benefits of the two vaccination programs are described as follows:

Program A: 200 lives are saved

Program B: 1/3 probability that 600 lives are saved

Tversky and Kahneman (1981, 1986) find that most people prefer Program A. Now, consider the following descriptions:

Program A: 400 die

Program B: 2/3 probability that 600 die

Kahneman and Tversky find that most people prefer Program B. (See also Baron, 2000.) Since the first set of descriptions is identical to the second, choices seem to be inconsistent.

a. How does this example relate to the framing problem discussed by Kahneman and Tversky?

b. How do the descriptions presented here pertain to risk-aversion and risk-seeking preferences?

10. Suppose that an investor has the opportunity (and funding ability) to pay $100,000 for a 50% chance to win $300,000 and a 50% chance of winning nothing.

a. What is the expected value of the gamble?

b. What is the standard deviation of payoffs for this gamble?

c. Suppose that you have the opportunity (and funding ability) to repeat participation in this gamble for a total of five gambles. Each wager's outcome is independent of

the outcomes of all other wagers (the correlation coefficient between wager payoffs is zero). What is the expected value of this set of five wagers?

d. What is the standard deviation for this set of five wagers?

e. Which set of wagers has a higher expected payoff—that described in parts a and b of this question or that described in parts c and d of this question?

f. Which set of wagers has a lower risk as measured by standard deviation—that described in parts a and b of this question or that described in parts c and d of this question?

g. Which set of wagers seems to be preferable based on your answers to parts a through f—the single wager or the set of five wagers?

h. Devise an argument that if an individual finds the gamble described in parts a and b unacceptable, he will also find the gambles described in parts c and d unacceptable.

CHAPTER

12

Market Efficiency

12.1 INTRODUCTION TO MARKET EFFICIENCY

An efficient capital market is defined as a market where security prices reflect all available information. More formally, one can state that in an efficient market, the expected price $E[\tilde{P}]$ of a tradable asset, given the information ϕ available to the market and the information ϕ_k available to any investor k, equals the expected price based on the information available to the market for all investors k:

$$E[\tilde{P}|\phi, \phi_k] = E[\tilde{P}|\phi] \; \forall k$$

Thus, the price of the asset reflects the consensus evaluation of the market based on the information available to the market, regardless of private information held by investor k. This statement implies that a given individual k's information set ϕ_k not shared with the market does not improve his estimate of expected price in an efficient market; the market price already reflects all relevant information ϕ including investor k's special information ϕ_k. In a perfectly efficient market where security prices fully reflect all available information, all security transactions will have zero net present value.

While we cannot realistically expect markets to be perfectly efficient, it is interesting to study the extent to which markets approach efficiency. For example, what type of information is reflected in security prices? What is the speed of information flow? How much time is required for a particular type of information to be reflected in security prices? Does market reaction to new information (or lack thereof) reflect rational self-serving behavior on the part of investors? If markets were perfectly efficient with regard to a given type of information, investors could not use this type of information to gain higher than normal returns; the information would already be reflected in security prices. Thus, investors might be able to earn abnormal returns if markets were slow to react to new information. Slow reactions to new information suggest that prices are predictable after new information is available.

If markets are perfectly efficient, then, by definition, securities reflect all available information. One might argue that security prices will fluctuate only as new information is realized by the market. New information should be expected to occur randomly; otherwise, it is not really new. Thus, one should expect prices in an efficient market to fluctuate

Financial Trading and Investing
DOI: https://doi.org/10.1016/B978-0-323-90955-6.00012-4

© 2023 Elsevier Inc. All rights reserved.

randomly. Samuelson (1965) formalizes this argument in his paper entitled "Proof that Properly Anticipated Prices Fluctuate Randomly." This means that many tests of market efficiency are concerned with the extent to which security prices fluctuate randomly. Furthermore, since prices in an efficient market properly reflect all that investors know about securities, all transactions, on an ex-ante basis, will have zero net present value. Investors pay for securities exactly what they are worth.

While one might argue that market efficiency is the necessary product of rationality and greed, and that market survival requires these characteristics of investors who determine price-setting, we already suggested (in Chapter 6, concerning the importance of noise traders) that perfect market efficiency would cause financial markets to fail. Furthermore, we will present arguments later that predictable prices might be perfectly consistent with market efficiency when investors are risk-averse and in certain other circumstances.

12.2 WEAK FORM EFFICIENCY AND TECHNICAL ANALYSIS

As discussed in the previous section, a capital market is efficient when security prices reflect all available information. One type of relevant information might be historical price sequences or trends. *Technical analysis* is concerned with finding situations where historical price sequences can be used to forecast price movements. *Weak form efficiency tests* are concerned with whether an investor might consistently earn higher than normal returns based on knowledge of historical price sequences and other trading information. One can never prove weak form efficiency because there are an infinite number of potential ways to forecast future returns from past returns. However, one might argue that certain tests imply efficiency (or inefficiency) with regard to a specific sequence or pattern of prices. An investor armed with the knowledge of a test indicating a market inefficiency might expect to earn a higher than normal return, or face a market impediment preventing him from doing so. On the other hand, any test of market inefficiency is actually a joint test for that particular inefficiency and of some model describing normal returns in an efficient market. Thus, market efficiency cannot be fully rejected unless one is certain that the correct model for normal returns has been selected for that market. The selected benchmark model might be some sort of risk-adjusted rate of return, perhaps based on CAPM, or might be as simple as the market return. Furthermore, Grossman and Stiglitz (1980) argue that the existence of costly information retrieval and processing must lead to abnormal return sequences to compensate analysts for bearing these costs. Thus, markets cannot be fully or perfectly efficient. In any case, some of the more significant weak form efficiency tests are described in the following paragraphs, following a brief introduction to technical analysis.

Technical Analysis: A Brief Introduction

Technical analysis is concerned with historical market price and volume sequences to evaluate or time securities transactions. Technical analysis is based on the concept that all information regarding securities, including earnings, risk, products, etc. is reflected in market behavior. Market price sequences are of primary importance; many technical

analysts focus on charting historical market prices of securities. Also important to many analysts are market volume and information specific to particular market segments.

The theoretical foundation for technical analysis derives from the following set of assumptions:

1. Market value is determined by the interaction of supply and demand, which are functions of a variety of factors, both rational and irrational.
2. Security prices tend to move in trends that persist for an appreciable length of time, despite minor fluctuations in the market. Many of these trends will repeat over time in a rather consistent manner.
3. Changes in a trend are caused by the shifts in supply and demand.
4. Shifts in supply and demand, regardless of why they occur, can be detected sooner or later in charts or sequences of market transactions.

Many technical analysts assume that security prices and market behavior are based on the "psychology" of the market. This psychology is revealed through historical price sequences and charts. A given sequence or pattern may be associated with a given market psychology, which will result in the same future price outcome as occurred when the same sequences or patterns occurred before. Thus, the market reacts the same way each time it encounters the same psychology as revealed by the price sequence. Much of the remainder of this section will be concerned with various technical rules and techniques and the extent to which markets efficiently capture historical price information.

Price Sequences, Momentum and Mean Reversion

Statisticians and economists with an interest in statistics have long been fascinated by the tremendous amounts of financial data made available by the various data services. For example, Cowles (1933), Working (1934), and Kendall (1953) studied weak form markets efficiency, or more specifically, the random movement of stock prices. Their results indicated that stock prices seemed to fluctuate randomly, without being influenced by their histories. Such processes are called random walks (see Chapter 7).

Another of the earlier weak form efficiency tests (Granger, 1968) found a very slight, but statistically significant relationship between historical and current prices: 0.057% of a given day's variation in the log of the price relative is explained by the prior day's in the log of the price relative:

$$\ln\left(\frac{P_{i,t}}{P_{i,t-1}}\right) = \hat{b}_0 + \hat{b}_1 \cdot \ln\left(\frac{P_{i,t-1}}{P_{i,t-2}}\right) + \hat{\varepsilon}_{i,t}$$

The r-square value from this regression was 0.00057, where $P_{i,t}$ represents price of stock i on a given day t, $P_{i,t-1}$ the price on the day immediately prior, $\hat{b}_0$ and $\hat{b}_1$ regression coefficients, and $\hat{\varepsilon}_{i,t}$ the error terms in the regression.[1] On the other hand, Fama and MacBeth (1973), after adjusting for risk, found no correlation in daily capital asset pricing model (CAPM) residuals.

[1]See text Appendix A.1 for a review of basic statistics topics. We will illustrate a mean reversion/momentum test in the Testing Momentum and Mean Reversion Strategies section of this chapter.

A typical testing format assumed that stock returns were produced by the CAPM (or single index model), but that residuals from such a model were uncorrelated with prior day residuals:

$$r_{i,t} = r_f + \beta_i(r_{m,t} - r_f) + \hat{\varepsilon}_{i,t}$$

$$\hat{\varepsilon}_{i,t} = \hat{a} + \hat{b}_i\hat{\varepsilon}_{i,t-1} + \hat{\eta}_{i,t}$$

where $r_{i,t}$ represents the return on stock i on a given day t, β_i its CAPM beta, and r_f the riskless return rate. Error terms (residuals or excess returns) $\hat{\varepsilon}_{i,t}$ are then regressed against their prior day values $\hat{\varepsilon}_{i,t-1}$. The sign of the $\hat{b}_i$ coefficient here would be key. A negative value for $\hat{b}_i$ suggests that sequential returns or price changes are inversely correlated, indicating that stock prices are more likely to drift towards their historical mean values than drift away. Positive values for $\hat{b}_i$ suggest that directions of sequential price changes are likely to be sustained, indicating that price changes might carry momentum. Fama and MacBeth found no (or very little) evidence for either mean reversion or momentum in stock prices.

Regardless of the analyst's findings concerning evidence for mean reversion or momentum in stock prices, it is important to note that correlation coefficients can be unduly influenced by extreme observations. This is because returns (or logs of one plus returns) and residuals are not normally distributed as required by CAPM and OLS regression assumptions. One way to deal with such assumption violations is to construct a simple runs test. For example consider the following daily price sequence: 50, 51, 52, 53, 52, 50, 45, 49, 54, and 53. The price changes might be represented by the following: (+ + + − − − + + −), indicating four price runs. That is, there were four series of positive or negative price change runs. The expected number of runs in a runs test if price changes are random is (MAX + MIN)/2, where MAX is the largest number of possible runs (equals the number of prices in the series minus 1) and MIN is the minimum number (1). The number of runs consistent with random sequences in our example is 5 = (9 + 1)/2. More runs than this number suggests price reversals (mean reversion), and a smaller number of runs suggests positive correlations between sequential price changes (momentum). Thus, our example might be consistent with slight evidence of momentum, but the sample set is too small to suggest significance. The actual levels of returns are unimportant; only the signs of returns are important, so that extreme observations will not unduly bias tests. In one test of daily price changes, Fama (1965) expected 760 runs based on the assumption that price changes were randomly generated, but only found 735 runs. This indicated a small positive correlation in daily stock price changes. Granger and Morgenstern (1972) found that high transactions costs seem to be related to runs—investors are unable to exploit a series because of brokerage commissions.

Neiderhoffer and Osborne (1966) found two to three times as many reversals of price trends as continuations based on transaction-to-transaction price data. This might be because of unexecuted limit orders—for them to be executed the price has to reverse itself. For example, suppose that a market purchase order has just been executed at an uptick. All of the limit sell orders at this most recent execution price have to be executed for the price to increase again. This means that a purchase is more likely to be followed by a downtick (−) or no change at all (0) than an uptick (+).

Relative Strength

Relative strength trading rules are concerned with the current price of a stock relative to a historical price or the prices of other securities or indices. For example, one relative strength rule compares the current price of a stock to its average price over a given period of time. Levy (1967), conducting a test of many such rules, found the following rule to be profitable:

1. Rank stocks based on the ratio (P_t/P_{avg}) where P_t is the current price and P_{avg} is the average price over the prior 27 weeks.
2. Buy the stocks in the highest fifth percentile and sell the stocks in the lowest 70th percentile.

This rule might suggest that the stocks with the strongest performance over the past 27 weeks will outperform the market. However, this rule might simply mean we buy the riskiest securities (high expected returns) and sell the lowest risk securities. Thus, perhaps, this rule simply means that investors purchasing higher risk stocks are merely compensated with higher returns for bearing this risk. Jensen and Bennington (1970) found that relative strength rules worked no better than random purchases of stocks after adjusting for transaction costs and even worse given risk adjustments. They suggest that the success of the rule suggested by Levy is due merely to selection bias, in that Levy had tested 68 different trading rules. That is, one of the 68 trading rules must surely work in his limited size sample of prices.

Filter Rules and Market Overreaction

A filter rule states that a transaction for a security should occur when its price has changed by a given percentage over a specified period of time. For example, Alexander (1961), in a study of stock prices from 1897 to 1959, suggested that certain price increases of given proportions indicated upward trends, and should serve as purchase signals. He concluded that there were identifiable trends in security prices, but that these trends varied over time. Other early studies (e.g., Fama and Blume, 1966) found that when filter rules did seem to work (however slightly), they were not able to cover transactions costs. Profitability of these rules seems to be related to daily correlations. Such correlation and filter rules seemed to work slightly better in Norway, where stronger correlations tended to exist. However, these markets were less liquid and transactions costs were significantly higher in Norwegian markets than in American markets (See, for example, Jennergren and Korsvold, 1975.). None of the tested rules seemed to work well enough to cover the particularly high transactions costs. Regardless, the apparent profitability of the filter rules seem directly related to transactions costs, and failed to cover them.

A few statistical studies suggest that market overreactions to information and abnormal returns can be realized by buying losers and selling winners. DeBondt and Thaler (1985) argued that buying stocks that performed poorly in a prior three- to five-year period and selling those that performed well would have generated abnormally high returns in subsequent three- to five-year periods. Apparently, longer-term stock-price mean reversion seems due to market overreactions over these longer periods. Richards (1997) found similar results for market indices in various countries.

On the other hand, Jegadeesh and Titman (1993) found results that conflicted with DeBondt and Thaler based on shorter holding periods (3 to 12 months). Their study suggested that the market is slow to react to firm-specific information. Lo and MacKinlay (1988) found that the level of short-term momentum effects displayed by stocks is inversely related to firm size. Apparently, smaller firms experience more momentum effects than do larger firms. If smaller firms have smaller investor followings and more restricted information availability, this apparent momentum might suggest that the market is relatively slow to react to information.

The findings of both DeBondt and Thaler and Jegadeesh and Titman that seem to contradict weak form market efficiency are not universally accepted. For example, Richardson and Stock (1988) argued that these momentum results of DeBondt and Thaler were due largely to their statistical methodology, as did Jegadeesh (1991), who argued that these mean reversion effects seemed to hold only in January.

Moving Averages

Moving average techniques consolidate shorter series of observations into longer series, are commonly used for smoothing data variability, and are frequently used as a reference point to gauge daily fluctuations. A simple $q-$period moving average is computed as follows:

$$MA_{q,t} = \frac{x_t + x_{t-1} + \cdots + x_{t-q-1}}{q} = \sum_{i=1}^{q} x_{t-i}/q$$

For example, daily stock prices are compared to a moving average of a specified number of historical prices. Trading strategies might be based on these moving averages. Perhaps, if current prices rise above a falling moving average, they might be expected to drop back towards the moving average; selling is suggested. The movement above the moving average might imply that the current high price is due to random error or noise. Buying is suggested when the moving average flattens out and the stock's price falls below the moving average. There are numerous other moving average rules, and they can be thoroughly contradictory with one another.

Moving averages can be computed for any number of price data points. For example, consider the following sequence of daily closing prices for a given stock over a period of time:

12	14	17	13	14	19	22	17	11	18	16	22
$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$	$t=7$	$t=8$	$t=9$	$t=10$	$t=11$	$t=12$

The following represents the sequence of simple three-day moving averages for the above price sequences:

NA	NA	14.3	14.7	14.7	15.3	18.3	19.3	16.7	15.3	15.0	18.7
$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$	$t=7$	$t=8$	$t=9$	$t=10$	$t=11$	$t=12$

A study by Brock, Lakonishok, and LeBaron (1992) demonstrated evidence suggesting that certain moving average rules and other rules based on resistance levels produced higher than normal returns when applied to daily data for the Dow Jones Industrial Average from 1897 to 1986. However, Sullivan, Timmerman, and White (1997) tested their findings on updated data and found "that the best technical trading rule does not provide superior performance when used to trade in the subsequent 10-year post-sample period."

The January Effect

Numerous studies including Rozeff and Kinney (1976), Roll (1981), Keim (1983), and Reinganum (1983) have confirmed a "January effect." This effect is that returns for the month of January tend quite consistently to exceed returns for any other of the eleven months. This January effect seems most pronounced during the first five trading days in January and the last trading day in December. However, the January effect continues to persist throughout the month. Additionally, this January effect seems to have a greater effect on the shares of smaller companies (which are frequently held by individual investors) than on the shares of larger firms (frequently held by institutional investors).

Roll (1983), Keim (1989), and Griffiths and White (1993) argued that much of the January effect can be explained by the tendency for December transactions to be seller initiated and execute at bid prices, while January transactions are buyer initiated and execute at offer prices. However, the latter two papers argue that the January effect is large enough that it would exist even if all transactions executed at bid prices. None of these papers explains the full size of the January effect.

The January Effect and Tax-Driven Selling

One explanation for this January effect is year-end tax selling—investors sell their "losers" at the end of the year to capture tax write-offs. This year-end tax selling forces down prices at the end of the year. They recover early in the following year, most significantly during the first five trading days in January (and the last trading day in December). Branch (1977) finds evidence suggesting this explanation. He found that stocks selling at their annual low at year's end outperformed the market in January by 8%. However, he concluded that this increase barely covered transaction costs for investors wishing to exploit this price increase (he assumed a high transaction cost). Nonetheless, an investor wishing to purchase securities in late December might benefit from this knowledge. Dyl (1977) supports the year-end selling hypothesis by reporting that abnormally high trading volume exists in December. Adding further credibility to the year-end tax selling hypothesis, other studies, including Givoly and Ovadia (1983), have also found that "losers" outperform "winners" in January of the subsequent year and that firm returns in January are directly related to losses in the prior year. Furthermore, Chan, Chen, and Hsieh (1985) and Chang and Pinegar (1986) report January effects for low-grade corporate bonds and in shares of companies that issue these bonds. This effect does not seem to hold for high-grade corporate bonds or for the shares of the companies that issue these bonds. This, they argue, supports the tax-loss selling hypothesis for the January effect.

Contrasting the tax explanations for the January effect are studies demonstrating that this effect exists in markets whose tax years differ from the calendar year. That is, if tax-loss selling explains the January effect, the January effect should not hold in countries whose tax years differ from calendar years. Nonetheless, the January effect does appear in Australia and other countries where the fiscal and calendar years differ. Furthermore, the January effect in Canada existed before the introduction of a capital gains tax. Why should a country experience a tax-related January effect if there is no tax-driven selling?

However, some observers have argued that U.S. markets are sufficiently influential in world markets that year-end tax selling in the U.S. might simply drive prices in other markets. On the other hand, studies have demonstrated a January effect in U.S. markets during the 1877–1916 period, prior to the introduction of U.S. income taxes—again, a January effect with no tax-driven selling. Furthermore, Kihn (1996) presents evidence that municipal bond issues, which are free from federal taxation, experience a significant January effect. Thus, despite the popularity of the tax explanation, there still remain serious doubts regarding its ability to explain the January effect.

The January Effect and Window Dressing

Some market observers have also suggested that funds will "window dress" at year-end by buying winners (stocks that performed well earlier in the year) and by selling losers. These transactions occur at the end of the year so that their clientele can see from year-end financial statements that their funds held high-performing stocks and did not hold losers. The buy pressure on winners and sell pressure on losers might be expected to cause losers to underperform winners during the period of window dressing; in January after the period of window dressing losers will outperform winners. However, most institutions will report their holdings to their clients more than once per year and some will report based on fiscal years other than from January 1 to December 31. But, this effect does not appear in any other month. Furthermore, winners still realize higher January returns than in any other month, just not as high as losers. If the "window-dressing" hypothesis explains the January effect better than the tax-selling hypothesis, one should expect that shares held by institutions should outperform shares held by individuals during the month of January. First, we note that the January effect is more pronounced for smaller firms than for larger firms (smaller firms are more likely to be held by individual investors). Second, another study found that the January effect is far more pronounced for companies with significant numbers of individual shareholders than those companies with significant numbers of institutional investors. Thus, the tax-selling hypothesis seems to explain the January effect better than the "window-dressing" hypothesis. However, since there are still many holes in the tax-selling hypothesis, it appears that the January effect remains an anomaly.

The Small Firm and Price/Earnings Effects

There also exists substantial evidence that smaller firms outperform larger firms (e.g., Banz, 1981). For example, if one were to rank all NYSE and Nasdaq firms by size, one is likely to find that those firms that are ranked as smaller will outperform those that

are ranked larger. This effect holds after adjusting for risk as measured by beta. However, other measures of risk may be more appropriate for smaller firms that may not have well-established trading records due to market thinness or due to "newness" when compared to typical larger firms. Furthermore, transactions costs for many smaller firms may exceed those for larger firms, particularly when they are thinly traded. There is also evidence that these abnormally high returns for smaller firms are most pronounced in January. Although Fama and French (1992) find a significant size effect in their study of the CAPM over a 50-year period, they do not find a size effect during the period between 1981 and 1990. This *might* suggest that the size effect either no longer exists or was merely a statistical artifact prior to 1981. Two studies found that abnormal returns in January are more related to share price effects than to firm size; in fact, after controlling for share prices (and noting the higher proportional transactions costs on lower priced shares), the size effect disappears as a partial explanation of the January effect.

Basu (1977) and Fama and French (1992) find that firms with low P/E ratios outperform firms with higher P/E ratios. Fama and French find that the P/E ratio, combined with firm size, predicts security returns significantly better than the CAPM.

The Initial Public Offering Anomaly

The initial public offering (IPO) anomaly refers to three unusual pricing patterns associated with IPOs of equities:

1. Short-term IPO returns are abnormally high.
2. IPOs seem to underperform the market in the long run.
3. Abnormal IPO return performance seems to be cyclical.

IPOs have been shown to generate significant short-term returns. These returns are most substantial on the day of the new issue. The abnormal returns exist after adjusting for risk (see, for example, Ibbotson and Jaffee, 1975; McDonald and Fisher, 1972; Ritter, 1984). Thus, it seems that unseasoned offerings with no trading histories do generate abnormally high short-term returns. However, the longer-term returns on IPOs do not seem nearly so high; in fact, several studies report IPO returns to be negative over the period 20 days to two years after the IPO (see, for example, Ritter, 1991). Some evidence suggests that these IPO returns are due to the provision of useful price-setting information by IPO market participants and by price supports in the IPO aftermarket. The third part of the IPO anomaly is that the first two are cyclical, being more pronounced during strong stock market periods.

Charting

Thus far, we have considered rather simple trading rules that can be rather easily put to statistical tests. Generally speaking, and with a few possible exceptions, there is little empirical evidence that most of these simple technical rules can cover transactions costs necessary to produce profits in stock markets.

Charting is the application of chart patterns in the analysis of security and other market prices. Charting allows the analyst to discern more complex trends than the simple rules we've discussed earlier in this chapter. While all chartists focus on prices and volume over time, they do make use of different types of charts. One of the most popular types of charts is the candlestick chart such as that in Figure 12.1. In most instances, a single candlestick figure will reflect the range of prices for a single stock, with the body of the candlestick in light green (gray in print version) (dark red (dark gray in print version)) depicting the security's open price at the bottom (top) of the body when the market increases (decreases) during the day. The shadows (vertical lines) extending above and below the candlestick body reflect the prices extending to the daily high at the top and low trading prices of the security.

The chartist will discern patterns depicted by the candlesticks. Such patterns might be relatively simple such as resistance levels or moving averages or might discern more complex patterns such as Elliott waves, head and shoulders, pennants or flags. Reading charts is often considered an art because their interpretation depends on the analyst's judgment on how to fit a huge variety of possible patterns over an infinity of possible time frames.

Two well-known charting patterns are the Head and Shoulders and the inverse (upside down) Head and Shoulders (see Figure 12.1) formations. Volume increases when the price reaches the top (bottom) of the head in the (inverted) Head and Shoulders formation strengthen the sell (buy) signal. The completion of the right shoulder when the neckline is broken is a sell (buy) signal in the (inverse) Head and Shoulders formation (as in Figure 12.1).



FIGURE 12.1 Candlestick chart, AAPL (Apple, Inc.), December 14, 2016. Inverse head and shoulders formation.

One of the better-known charting tools is based on the *Dow Theory*, first suggested by Charles Dow in the late 1800s, which examines trends in the market and classifies them into three basic types:

1. Primary Trends: Commonly called bear or bull markets, these trends represent the main component of the Dow Theory. They are long term, typically 4 years in duration.
2. Secondary Movements: Referred to as corrections, these movements last only a few months.
3. Tertiary Moves: Daily fluctuations that are considered random variations. Although these movements are not essential to the theory, they are plotted to delineate primary and secondary trends.

A line chart may be plotted using each day's closing (or opening, or high, or low) price, and connecting them with a line. Short-term trends may only represent a secondary movement, however a long term trend, with each secondary trend failing to reach a new bottom may be considered a bull market.

There are a huge variety of such charting theories and formations. In some respects, charting is like astrology. Its validity is impossible to disprove, and many of its principles are difficult to justify. However, it is used by many traders and investors, particularly in currency and commodities markets, and particularly for short-term trading executions. Understanding how these traders and investors behave may be important to understanding and profiting from stock market behavior.

Sports Betting Markets

Sports betting markets potentially have much in common with stock markets. Sports betting markets resemble financial markets in that both have large numbers of greedy participants who risk capital, have varying levels and types of knowledge and have strong financial incentives to quickly and rationally act on the best information in competitive settings. Thus, we might expect a reasonable amount of efficiency to new information in both markets. However, there is some evidence of persistent weak-form inefficiencies in sports betting markets. For example, Thaler and Ziemba (1988) and numerous other studies note that favorites in horse races outperform long shots, though Woodland and Woodland (1994) find the opposite is true for baseball betting. Brown and Sauer (1993) find that several observable variables in addition to the spread can be used to improve the outcomes of professional basketball games and that teams with winning streaks systematically receive point spreads in subsequent games that are better than they actually experience. Gray and Gray (1997), Golec and Tamarkin (1991), Sauer, Brajer, Ferris, and Marr (1988), and Gandar, Zuber, O'Brien, and Russo (1988) find evidence that certain strategies can be used to improve professional football betting. Consider these rhetorical question: Might observed inefficiencies and biases in sports betting markets be similar in frequency and nature (e.g., momentum and mean reversion) to those in financial markets? Do these inefficiencies reflect profit-making opportunities?

Summary

Generally, statistical studies seem to indicate that stock markets are efficient with respect to historical price sequences. However, one must realize that a huge number of possible sequences can be identified with any series of prices. Clearly, many of these series must be associated with higher than normal future returns, but will the associations continue into the future? Statistical studies have certainly found evidence of predictability in stock prices, normally referring to such findings as anomalies. However, when research finds a sequence that leads to higher than normal returns, one must question whether the abnormal return result is merely a statistical artifact due to data mining. According to William Schwert (1983):

> These [research] findings raise the possibility that anomalies are more apparent than real. The notoriety associated with the findings of unusual evidence tempts authors to further investigate puzzling anomalies and later try to explain them. But even if the anomalies existed in the sample period in which they were first identified, the activities of practitioners who implement strategies to take advantage of anomalous behavior can cause the anomalies to disappear (as research findings cause the market to become more efficient).

Richard Roll (1992), in a blunt comment, stated:

> I have personally tried to invest money, my client's and my own, in every single anomaly and predictive result that academics have dreamed up. That includes the strategy of DeBondt and Thaler (that is, sell short individual stocks immediately after one-day increases of more than 5%), the reverse of DeBondt and Thaler which is Jegadeesh and Titman (buy individual stocks after they have decreased by 5%), etc. I have attempted to exploit the so-called year-end anomalies and a whole variety of strategies supposedly documented by academic research. And I have yet to make a nickel on any of these supposed market inefficiencies.

Clearly, technical analysis has its share of critics. For example, Warren Buffet was quoted saying, "I realized technical analysis didn't work when I turned the charts upside down and didn't get a different answer." In any case, most apparent incidences of mispricing seem eliminated by transactions costs. The primary exceptions to weak form market efficiency seem to be the IPO effects, probably the January effect, perhaps the small firm effect, and perhaps the P/E effect. There seems to be little agreement as to why these effects persist or even whether the latter two even exist; they remain anomalies.

12.3 TESTING MOMENTUM AND MEAN REVERSION STRATEGIES

In this section, we provide an illustration of a momentum/mean reversion test. Essentially, this means that we will test whether the return for a stock on a given day can be used to predict the return on that stock for the subsequent day. (See the discussion on price sequences and relative strength in the previous section.) If returns seem to be positively correlated from one day to the next, this would evidence momentum in stock prices over time. A negative correlation would evidence mean reversion. Consider the sequence of daily closing prices in Table 12.1 for a hypothetical stock Q, from which we compute

TABLE 12.1 Stock Price and Return Data

Date	t	Price_t	Return_t	Return_{t-1}
02/01/12	1	49		
02/02/12	2	50	0.020408	
02/03/12	3	51	0.020000	0.020408
02/04/12	4	52	0.019607	0.020000
02/05/12	5	55	0.057692	0.019607
02/06/12	6	57	0.036363	0.057692
02/07/12	7	58	0.017543	0.036363
02/08/12	8	59	0.017241	0.017543
02/09/12	9	58	−0.016949	0.017241
02/10/12	10	55	−0.051724	−0.016949
02/11/12	11	53	−0.036363	−0.051724
02/12/12	12	52	−0.018867	−0.036363

returns. The prices given in this table assume that the stock traded each day, including weekends. We will attempt our test illustration based on an extremely small sample to simplify our computations.

Although there are many ways to determine the nature of the relationship between the return on a security and its prior day's return, we will examine whether there exists a linear relationship based on a simple ordinary least squares (OLS) regression of the form: $\hat{r}_{i,t} = \hat{a} + \hat{b}_i \hat{r}_{i,t-1} + \hat{\varepsilon}_{i,t}$. We report regression results as follows:

$$\hat{r}_{i,t} = -0.0020142 + 0.7717758 \cdot \hat{r}_{i,t-1} \qquad r^2 = 0.57$$
$$(-0.25959)\ (3.259325) \qquad \text{SSE} = 0.00450311$$
$$n = 10 \quad d.f. = 8$$

Note that t-statistics are given in parentheses. The slope term $\hat{b}_i$ is 0.7717758, positive and statistically significant at the 5% level. We will return shortly to explain and interpret these results. However, we will first work through the details on obtaining the results of our simple OLS regression. First, our $\hat{a}_i$ and $\hat{b}_i$ coefficients are determined following the simple OLS method as follows:

$$\hat{b}_i = \frac{\sum (r_t - \overline{r}_t)(r_{t-1} - \overline{r}_{t-1})}{\sum (r_{t-1} - \overline{r}_{t-1})^2} = \frac{0.000775}{0.001004} = 0.7717758$$

$$\hat{a}_i = \overline{r}_t - \hat{b}_i \overline{r}_{i,t-1} = -0.0020142$$

To obtain regression diagnostics, we compute squared error-term values as in Table 12.2.

TABLE 12.2 Stock Price Data

t	Return_t	Return_{t-1}	$a + br_{t-1}$	ε_i	ε_i^2
3	0.020000	0.020408	0.013735	0.00626	0.000039239
4	0.019607	0.020000	0.013420	0.00618	0.000038278
5	0.057692	0.019607	0.013118	0.04457	0.001986848
6	0.036363	0.057692	0.042510	−0.00614	0.000037788
7	0.017543	0.036363	0.026049	−0.00850	0.000072353
8	0.017241	0.017543	0.011525	0.00571	0.000032673
9	−0.016949	0.017241	0.011291	−0.02824	0.000797554
10	−0.051724	−0.016949	−0.015095	−0.03662	0.001341651
11	−0.036363	−0.051724	−0.041934	0.00557	0.000031029
12	−0.018867	−0.036363	−0.030079	0.01121	0.000125692
				SSE =	0.004503110

We are able to compute regression standard errors from calculations in Table 12.2:

$$se(\hat{b}_i) = \sqrt{\frac{\frac{SSE}{n-2}}{\sum (r_t - \overline{r}_t)^2}} = \sqrt{\frac{0.000562888}{0.01003}} = \sqrt{0.056069} = 0.23679$$

$$se(\hat{a}_i) = \sqrt{\frac{\frac{\sum_{i=1}^{n} \varepsilon_i^2}{n-2} \cdot \sum_{i=1}^{n} x_i^2}{n \cdot \sum_{i=1}^{n} (x_i - \overline{x})^2}} = \sqrt{\frac{0.000562888 \cdot 0.0107417}{10 \cdot 0.01003911}} = \sqrt{0.000060228} = 0.007760684$$

Given our standard error estimates $se(\hat{a}_i) = 0.007760684$ and $se(\hat{b}_i) = 0.05606069533$, we find our t-statistics to be $t(\hat{a}_i) = -0.25959$ and $t(\hat{b}_i)$ to be 3.2593. Suppose our tests were expressed formally as follows:

$$H_0: \hat{a}_i = 0 \quad H_0: \hat{b}_i = 0$$

$$H_A: \hat{a}_i \neq 0 \quad H_A: \hat{b}_i \neq 0$$

Since we will assume that a and b follow student t-distributions, we will compare our t-statistic to critical values found in the t-distribution table (See Table A.2 in the Text Appendix.), given the appropriate level of confidence and degrees of freedom. We are performing two-tail tests for significance. Suppose we test for significance at a level of 0.95 (0.025 on each side of the tail). In this case, the significance of our test is $\alpha = 0.025$ and the number of degrees of freedom k equals $10 - 2 = 8$. Our critical value for the test is found:

$$t^8_{0.025} = 2.306$$

While we fail to reject our first null hypothesis that $\hat{a}_i = 0$ (not an important finding here), we reject our second null hypothesis that $\hat{b}_i = 0$. Thus, our test does provide evidence that the returns for this security are linearly related to returns on prior days, and for a momentum effect in stock prices.

12.4 SEMISTRONG FORM EFFICIENCY

Investors acquire information for input to their valuation models and for security selections. If markets are efficient, their valuation results will reflect all available information. *Semistrong form efficiency* tests are concerned with whether security prices reflect all publicly available information. For example, how much time is required for a given type of information to be reflected in security prices? What types of publicly available information might an investor use to generate higher than normal returns? The majority of studies of semistrong form market efficiency suggest that the tested publicly available information and announcements cannot be used by the typical investor to secure significantly higher than normal returns. A few of the exceptions to this rule are included in the following paragraphs. In addition, investors able to react within a few minutes to event news may be able to secure higher than normal returns.

Early Tests

Garfield Cox (1930) found no evidence that professional stock analysts could outperform the market. Similarly, and more rigorously, Cowles (1933) performed several tests of what was later to be known as the efficient market hypothesis (EMH). He examined the forecasting abilities of 45 professional securities analysis agencies (including fire insurance companies, financial services companies, and financial publications). He compared the returns that might have been generated by professionals' recommendations to actual returns on the market over the same period. He found that the average returns generated by professionals were less than those generated by the market over the same periods. He found that the best performing fund was not an outlier; that is, it did not exhibit unusually high performance at a statistically significant level. Cowles also tested whether analyst recommendations were correct an unusually high number of times; that is, he tested whether analyst picks were profitable relative to the market more frequently than might be expected with random picks. Analyst picks were not more profitable.

Cowles also examined the abilities of analysts to predict the direction of the market as opposed to selecting individual stocks. He found that a buy and hold strategy was at least as profitable as following "average" advice of professionals as to when to be long or short in the market. He performed a simulation study using a deck of cards (since there were no computers capable of generating random numbers at the time). Based on reports of analyst recommendations, he computed the average number of times analysts change their recommendations over a year (33 times). He then randomly selected 33 dates, using cards numbered 1 to 229 (the number of weeks the study covered) to make simulated random recommendations. Draws were taken from a second set of randomly selected cards

numbered 1 to 9, each with a certain recommendation (long, short, half stock, and half cash, etc.) for a given date. Cowles then compared the results distribution of the 33 recommendations based on randomly generated advice to the advice provided by the actual advisors. He found that the professionals generated the same return distributions as did the random recommendations. Thus, he concluded that the best-informed analysts would perform no better than the uninformed investor making random investing decisions. Cowles also examined 255 editorials by William Peter Hamilton, the fourth editor of the *Wall Street Journal*, who had gained a reputation for successful forecasting. Between 1902 until his death in 1929, Hamilton forecast 90 changes in the market; 45 were correct and 45 were incorrect.

If experts are unable to distinguish between strong and weak stock market performers, and investors are well aware of this lack of ability, why do market forecasters still exist and investors still purchase and follow their advice? One possible explanation for reliance on unreliable "expert" forecasters is that investors are less interested in accuracy than in avoiding responsibility for their selections. Investors who rely on advice from experts seek to avoid blame when the forecasts are inaccurate. Avoidance of responsibility in another field is illustrated by Steadman and Cocozza (1978) in their study of New York psychiatrists who were asked to predict whether mental patients were dangerous and required involuntary confinement. Although numerous studies have consistently shown that psychiatrists cannot predict which patients are dangerous, the expert diagnoses were accepted by 87% of the courts in their study. Such "expert" advice seemed to relieve courts of further responsibility for the patients.

Stock Splits

In a particularly important seminal test of semistrong form market efficiency, Fama, Fisher, Jensen, and Roll (1969) examined the effects of stock splits on stock prices. Because it seems logical that stock splits should be cosmetic in nature, and that Fama et al. generally reached this empirical conclusion, the results of this paper are somewhat less important than the methodology used in the study. This paper was the first to use the now classic event study methodology. Although stock prices did change significantly before announcements of stock splits (and afterwards as well), Fama et al. argued that splits were related to more fundamental factors (such as dividends), and that it was actually these fundamental factors that affected stock prices. The splits themselves were unimportant with respect to subsequent returns.

Fama et al. identified the month in which a particular stock split occurred, calling that event month time zero for that stock. Thus, each stock had associated with it a particular month zero ($t = 0$), and months subsequent to the split were assigned positive values. They then estimated expected returns for each month t for the stocks in their sample with single index model: $r_{i,t} = a + b_i r_{m,t} + e_{i,t}$ where the expected residual ($e_{i,t}$) value was zero. Fama et al. tested 940 splits occurring between from 1956 to 1960, excluding from their beta computations returns data 15 months before and after splits. They then examined residuals ($e_{i,t}$) for each security i for each month t then averaged the residuals (AR_t) for each month across securities. Afterwards, they calculated cumulative average residuals (CAR_t) starting

30 months before splits ($t = -30$). Cumulative average residuals increased dramatically starting 30 months before split. Fama et al. regarded it unlikely for this increase to occur because a split was anticipated, but reasoned that splits occur after significant price growth. They found that after splits, residuals (AR_t) again average zero. Finally, Fama et al. split their sample of companies into those increasing dividends after a split versus companies not increasing dividends. Companies splitting stock and then increasing dividends had continued increasing CARs after the split announcement date; those splitting stock then decreasing dividends experienced decreasing CARs. Thus, dividends might indicate fundamental strengths; splits do not appear to be relevant. On average, once the split is announced, positive residuals (CARs) no longer exceed zero.

Subsequent tests on stock splits have not been entirely consistent with the results of Fama et al. For example, it has been argued that splits increase the proportional trading costs of stocks. Investors should require higher returns to compensate for these higher trading costs. Later studies have documented positive residuals on split announcements.

Nonetheless, the Fama et al. study provided the framework for future event studies and semistrong efficiency tests. Consider the following general notes regarding testing the semistrong form efficiency hypothesis:

1. Use daily price and returns data since information is incorporated into prices within days (or much shorter periods).
2. Announcements are usually more important than events themselves.
3. Base security performance on estimated expected returns.
4. When using the market model (standard single index model), we estimate slopes from historical data. Normally, we find them biased forecasters for future values, so we may adjust them towards one.
5. One way to deal with slope measurement error is to use moving windows for the period whose excess return is being determined, estimate slope based on time periods preceding and following the testing period, excluding the testing period itself.
6. An alternative to adding to determine cumulative excess returns is adding them to 1, then multiplying them to compute *buy and hold abnormal residuals* as follows: $BHAR_T = \Pi(1 + e_t) - 1$.

Corporate Merger Announcements, Annual Reports, and Other Financial Statements

Thousands of other tests of semistrong form efficiency have been reported in the academic literature, covering wide varieties of events. For example, Firth (1976) considered market efficiency when an announcement is made for purchase of more than 10% of a firm. Presumably, an announcement indicates a potential merger. Firth calculated CAR starting 30 days prior to announcements; the bulk of CAR is realized between last trade before and first trade after announcements, though it still increases slightly after an announcement. Thus, a large block purchaser can still make excess returns. An insider obviously can make excess returns; one without inside information cannot (except very shortly after the announcement). Since returns change almost immediately, Firth suggested that there is semistrong efficiency with respect to merger announcements.

Ball and Brown (1968) study the usefulness of the information content of annual reports. With a primary focus on EPS, they find that security prices already reflect 85% to 90% of information contained in annual reports; security prices show no consistent reactions to annual report releases. They conclude that analysts obtain more timely information from other sources.

Information Contained in Publications and Analyst Reports

Davies and Canes (1978) considered information analysts sell to clients then publish in the "Heard on the Street" column in *The Wall Street Journal*. They used the market model to measure the relationship among the market, risk, and the security. Information in this column was frequently sold by investment firms to clients before publication in the journal. Prices seemed to rise significantly after information was sold to clients, then even more when it was published in the *Wall Street Journal*. Davies and Canes tested to see whether these large residuals on the *Wall Street Journal* publication day were significant by standardizing each day's return and then checking to see how many standard deviations from zero the excess or abnormal return lies.

Other studies (e.g., Black, 1973; Holloway, 1981; Huberman & Kandel, 1987) have been performed on the ability to use rankings and other information provided by *Value Line Investment Surveys* to generate profits. Although they are not consistent, many studies, particularly those before 1990, seem to suggest that *Value Line* reports can be used to generate higher than normal returns. However, several later studies suggest that the excess returns based on *Value Line* analyses may not have been sufficient to cover trading costs or may have been due to systematic risk, momentum effects, and other factors (e.g., Nayar, Singh, and Yu, 2011).

More general studies on the value of analyst reports are somewhat mixed. The earlier study by Cowles (1933) found no evidence of value in analyst reports. On the other hand, Green (2006) found in his study of 7000 recommendation changes from 16 brokerage firms from 1999 to 2002 that, after controlling for transaction costs, purchasing (selling) quickly following upgrades (downgrades) resulted in average two-day returns of 1.02% (1.50%). He found that short-term profit opportunities persisted for two hours following the premarket release of new recommendations.

Womack (1996) found that analysts' mean postevent drift averages 2.4% on buy recommendations and is short-lived. However, sell recommendations result in average losses of 9.1% that are longer lived. These price reactions seem more significant for small-capitalization firms than for larger capitalization firms. Also, consider that sell recommendations may be particularly costly to brokerage firms, potentially damaging investment banking relationships and curtailing access to information in the future. Clearly, buy recommendations far outnumber sell recommendations and an incorrect sell recommendation may be particularly damaging to an analyst's reputation.

Analyst Reports and Conflicts of Interest

Michaely and Womack (1999) studied after-market returns of 391 firms going public in 1990 and 1991, attempting to discern whether analysts working for firms underwriting the IPOs provided buy recommendations that were superior to those of investment

institutions not participating in the underwriting efforts. Results suggest that if the analysts worked for an institution that did not participate in the underwriting, they were more likely to recommend a stock that had performed well in the recent past and would continue its strong performance. However, if the analysts worked for a firm that participated in bringing the IPO to the market, it was more likely to have recorded poor performance both before and after the analysts' recommendation. This evidence suggests that analysts working for investment banks are likely to attempt to prop up the prices of their underwritten securities with their recommendations.

In response to these apparently biased and unethical analyst recommendations, the Securities and Exchange Commission (SEC) announced in 2003 the *Global Research Analyst Settlement* with 10 of the industry's largest investment banks. This settlement resulted from investigations by Congress, the Office of New York Attorney General Elliot Spitzer, the SEC, and other regulators into apparent conflicts of interest among security analysts working for investment banks. The settlement required the 10 investment banks to pay $875 million in penalties and profit disgorgement, $80 million for investor education, and $432.5 million to fund independent research. In addition to these payments, the investment banks were required to separate their investment banking and research departments and add certain disclosures to their research reports. Nevertheless, Barber, Lehavy, and Truman (2007) found that between February 1996 and June 2003, buy recommendations of independent research firms outperformed those of investment banks by an average of 3.1 basis points per day. Investment bank hold/sell recommendations, in contrast, outperformed those of the independent research firms by an average of 1.8 basis points daily. Thus, investment bank buy opinions still underperform those of independent analysts, despite their other recommendations outperforming those of their independent competitors.

Discounted Cash Flow Analysis and Price Multiples

In their study of 51 highly leveraged transactions (management buyouts and leveraged recapitalizations), Kaplan and Ruback (1995) found that discounted cash flow (DCF) analysis provided better estimates of value than did price-based multiples. The free cash flow estimates used by Kaplan and Ruback were provided by management to the public as required by SEC regulation in going private transactions. Discount rates were estimated with CAPM (and other methods were used as well). Kaplan and Ruback found that between 95% and 97% of firm value was explained by (as indicated by r-square) DCF and slightly less was explained by price-based multiples. The price-based multiples did add useful information to the valuation process. In addition, analysts may be able to improve on DCF estimates by refining those provided by management. Regardless, the high r-squares do suggest that reasonably reliable information and analytical techniques are available in the marketplace for analysts and investors to use in company valuations.

Political Intelligence Units

Investors with money at stake have obvious incentives to access and quickly exploit information. Many investors and institutions are able to access and exploit important

information before it can be gathered and disseminated by the news agencies. Consider the case of USG Corporation, whose shares increased by 5.4% over two days prior to November 16, 2005 when Senate Republican Majority Leader Bill Frist announced that there would be a full Senate vote on a bill to create a $140 billion public trust for asbestos liability claims (see Mullins and Scannell, 2006). This fund would pay medical expenses and resolve lawsuits involving thousands of cancer victims who blamed USG, W.R. Grace, and Crown for their illnesses. Share prices of all these firms increased over the two days prior to November 16. However, the price increases clearly preceded the trust fund announcement. Returns for these firms over the two-day period exceeded those of the market. In addition, returns experienced by these particular firms far exceeded returns of their peer firms that were not involved in asbestos litigation. On the date that the actual announcement was finally made, these three firms showed no substantial reaction.

The SEC initiated an informal investigation to determine whether and how information might have been leaked to investors prior to its announcement. While staff members for Senator Frist claim to have been careful not to leak information prior to the announcement, the bill's authors, Senators Spector and Leahy, had held extensive discussions with lobbyists. Several law firms, including Sonnenschein Nath & Rosenthal, LLP and DLA Piper, have operated "political intelligence" units enabling their clients to obtain public policy information from lobbyists operating in Washington. These firms and political intelligence units include hedge funds as clients. Several hedge funds holding substantial stakes in affected companies belonged to the Financial Institutions for Asbestos Reform, an industry advocacy group, giving them additional opportunities to access information provided by lobbyists.

While it is not always clear whether any laws have been broken by such activity, it does appear that hedge funds may have successfully gained an information edge in their trading. Nevertheless, the 2012 passage of the Stop Trading on Congressional Knowledge (STOCK) Act restricted the insider trading activity and "tippee" potential of members of the legislative, executive, and judicial branches of government along with their staffs. Gao and Huang (2016) found that hedge funds with strong connections to lobbyists significantly outperformed index benchmarks prior to 2012, and that this out-performance was diminished after 2012.

Market Volatility

If security price changes are purely a function of information arrival, then security price volatility should be the same when markets are closed as when they are open. For example, stock return variances should be three times as high over a weekend as over a 24-hour period during weekdays. However, Fama (1965) and French (1980) found that return variances were only around 20% higher during weekends. On the other hand, one might argue that the arrival of new information over weekends is slower. But another study by French and Roll (1986) found that agricultural commodity futures prices (orange juice concentrate) were substantially more volatile during trading days than during weekends. Agricultural commodity futures prices are primarily a function of weather, news about which occurs over the weekend just as efficiently as during trading days. This suggests

that volatility, to some extent, is caused by trading itself. More generally, Shiller (1981) found that the volatility of stock prices significantly exceeded the level that might be explained by the volatility of dividends.

12.5 THE EVENT STUDY METHODOLOGY

An event study is concerned with the impact of a given firm-specific corporate event on the prices of the company's securities. For example, an event study might be conducted for the purpose of determining the impact of corporate earnings announcements on the stock price of the company. Event studies are used to measure market efficiency and to determine the impact of a given event on security prices. More important, from a trading perspective, event studies are used to back-test price data to determine the usefulness and reliability of trading strategies.

A number of studies have suggested a relatively high degree of efficiency in capital markets. If this suggestion is true, then one would expect that security prices would continuously reflect all available information. If security prices are a function of all available information, and new information occurs randomly (otherwise, it would not be new information), then one would expect that security prices would fluctuate randomly as randomly generated news is impounded in security prices. Thus, the "purchase or sale of any security at the prevailing market price represents a zero NPV transaction" (Brown & Warner, 1980). In a perfectly efficient market, any piece of relevant new information would be immediately reflected in security prices. One should be able to determine the relevance of a given type of information by examining the effect of its occurrence on security prices. Thus, nonrandom performance of security prices immediately after a given event suggests that news of the event has a significant effect on security values. The degree of efficiency in a market to a given type of information may be reflected in the speed that the market reacts to the new information.

At any given point in time, security prices might be affected by a large number of randomly generated pieces of new information or events. An event study is concerned with the impact of a specific type of new information of a security's price. Given that more than one piece of news may be affecting the security's price at any point in time, one will probably need to study more than one firm to determine how the given type of information will affect securities. Thus, a population or sampling of firms experiencing the given event will be gathered; the impact of the event on each of the firms' securities will be studied simultaneously. Thus, the first step in conducting an event study is to gather an appropriate sample of firms experiencing the event.

The impact of the event on security prices is typically measured as a function of the amount of time that elapses between event occurrence and stock price change. In a relatively efficient market, one might expect that the effect of the event on security prices will occur very quickly after the first investors learn of the event. Event studies are usually based on daily, hourly, or even trade-to-trade stock price fluctuations. However, we frequently are forced to study only daily security price reactions when more frequent data are not readily available.

In addition, if markets are relatively efficient, one should obtain security price information as soon as possible after the event is known, although determining when the information is known may be problematic. For example, analysts are often able to predict with a reasonably high degree of accuracy the earnings of firms and trade securities on the basis of their predictions (See, for example, Firth, 1976; Jaffee, Keim, & Westerfield, 1989.). Thus, the impact of corporate earnings changes may be realized in security prices long before earnings reports are officially released. Consequently, one may need to study the impact of a given event, news item, or announcement by considering security price reactions even before the event occurs. One should also take care in deciding on the precise nature of the event. For example, a dividend announcement may be of much greater interest than actual payment of the dividend.

Event studies typically standardize security price reactions by measuring the timing of security price reactions relative to the date of the event. For example, suppose that Company X announced its earnings on January 15 and Company Y announced its earnings on February 15. Let the base period time ($t = 0$) for Company X be January 15 and the base period time ($t = 0$) for Company Y be February 15. January 16 and February 16 (one day after the events) will be denoted as ($t = 1$) for the respective companies. Thus, the timing of the corporate events are standardized and we are able to measure average security price reactions 1, 2, etc., days after (and before) the event occurs.

Although stock return generating processes might be modeled as a random walk if capital markets are efficient, one should expect a long-term general drift in returns; that is, one might expect that investors will earn a "normal" return on their securities. Thus, excess or abnormal return randomness is observed when markets are efficient, except for a very short period after relevant new information is available. The abnormal return or residual in a given period for security i, $\varepsilon_{i,t}$, is the difference between its total, actual, or ex-post return $r_{i,t}$ and its expected, normal, or ex-ante return $E[r_{i,t}]$: $\varepsilon_{i,t} = r_{i,t} - E[r_{i,t}]$. To measure the impact of an event on security returns, one must have a consistent means of measuring normal returns. Brown and Warner (1980, 1985), in their classic studies of event study methodologies, suggest three models of normal returns:

1. *Mean Adjusted Returns*: The normal return for a security i equals a constant K_i. Typically, the mean return for the security over a sampling of time periods outside of the testing period serves as the constant K_i. The expected return for the security is assumed to be constant over time, though ex-ante returns will vary among securities. Thus, the abnormal return for the security is found: $\varepsilon_{i,t} = r_{i,t} - K_i$.
2. *Market Adjusted Returns*: The normal return for a security at a given point in time equals the market return for that period. The expected returns for all securities are assumed to be the same during a given period, though they vary over time. Abnormal returns are found: $\varepsilon_{i,t} = R_{i,t} - R_{m,t}$.
3. *Market and Risk Adjusted Returns*: Here, normal returns are assumed to be generated by a single index model. Typically, security returns are linearly related to market returns through stock betas. These risk-adjusted returns vary across securities and over time. Abnormal returns can be determined: $\varepsilon_{i,t} = R_{i,t} - \beta_i(R_{m,t} - r_{f,t})$.

One might test the significance of an event by averaging the abnormal performance for the sampling of securities around the event dates. If abnormal returns are not statistically

significantly different from zero during the relevant testing period, one may conclude that the test did not provide evidence indicating the significance of the event. If either no abnormal performance is detected around the event date or abnormal performance rapidly disappears, we have evidence of market efficiency with respect to that type of information.

Illustration: Event Studies and Takeovers

Numerous firms have been targets of takeover attempts in recent years. There is good reason to expect that targets of takeover attempts experience significant positive abnormal returns. This suggests that investors with prior knowledge (or strong suspicion) of an impending takeover can realize abnormal returns by purchasing shares of the target firm prior to the takeover announcement. Is this true? We will conduct our own event study to test the following hypotheses:

1. Takeover targets experience positive stock price reactions to takeover announcements.
2. Markets react efficiently to takeover announcements.

The event that we will study is takeover announcements. We need to locate an appropriate sampling of companies to study. Suppose we wish to base our study on the three targets of takeover attempts listed in Table 12.3.[2]

Obviously, an actual statistical study would certainly require a much larger sampling of firms, but for sake of computational ease, we will focus on only three firms here. Suppose that we establish a 21-day testing period (event window) for returns around the event dates, the event date plus 10 days before and 10 days after. Table 12.4 provides our three target firm stock prices during 21-day periods around merger announcement dates. We standardize event dates (merger announcements occur on the 11th date, standardized at day 0) and compute returns for each stock during each of the days in the testing period as in Table 12.5. The return for a particular day is simply the closing price for that day divided by the closing price for the prior day minus one: $(P_t/P_{t-1}) - 1$.

The next step in this study is to determine normal or expected returns for each of the securities for each date. We could use any of the three adjustment methods discussed above (with more information). Suppose that we have decided to use the Market Adjusted Return method. In this case, we compute daily returns for the market index for each day in our 21-day testing period for each stock. Table 12.6 lists adjusted S&P index values for each of the 21 dates affecting each of the three stocks along with returns for the index for each of those dates.

Next, based on actual returns computed in Table 12.5 and normal returns from the S&P 500 Index from Table 12.6, we compute periodic residuals $\varepsilon_{i,t}$ (abnormal returns) for each stock during each date in the testing period along with the average residual over the sample for each date as in Table 12.7.

One of our objectives is to determine whether any daily residual is statistically significantly different from zero. Following standard hypotheses testing techniques reviewed in

[2] A sample size of three for our study is absurdly small to be able to reach meaningful conclusions. We will examine only three companies merely for ease in computations and illustration.

TABLE 12.3 Takeover Targets

Target Firm	Symbol	Announcement Date
Fleet Boston	FBF	10/27/2003
Disney	DIS	02/11/2004
AT&T Wireless	AWE	01/18/2004

Note: The successful takeover of Fleet Boston by Bank of America was consummated in 2004. Comcast's bid for Disney failed. AT&T Wireless, the target of competing offers from Vodafone and Cingular, was acquired by Cingular in 2004.

TABLE 12.4 Target Company Stock Prices

Day	Calendar Date	FBF Price	Calendar Date	DIS Price	Calendar Date	AWE Price
10	10–Nov–03	39.69	26–Feb–04	26.73	3–Feb–04	11.08
9	7–Nov–03	39.84	25–Feb–04	26.30	2–Feb–04	11.16
8	6–Nov–03	40.26	24–Feb–04	25.96	30–Jan–04	11.05
7	5–Nov–03	39.89	23–Feb–04	26.75	29–Jan–04	11.03
6	4–Nov–03	39.60	20–Feb–04	26.55	28–Jan–04	11.02
5	3–Nov–03	39.86	19–Feb–04	27.00	27–Jan–04	11.17
4	31–Oct–03	39.73	18–Feb–04	26.71	26–Jan–04	10.93
3	30–Oct–03	39.45	17–Feb–04	26.90	23–Jan–04	10.61
2	29–Oct–03	38.90	13–Feb–04	26.92	22–Jan–04	10.56
1	28–Oct–03	38.17	12–Feb–04	28.00	21–Jan–04	10.99
0	27–Oct–03	38.56	11–Feb–04	27.60	20–Jan–04	10.39
−1	24–Oct–03	31.28	10–Feb–04	24.08	16–Jan–04	9.99
−2	23–Oct–03	31.54	9–Feb–04	23.77	15–Jan–04	9.81
−3	22–Oct–03	31.47	6–Feb–04	23.35	14–Jan–04	9.99
−4	21–Oct–03	31.88	5–Feb–04	23.20	13–Jan–04	8.55
−5	20–Oct–03	32.17	4–Feb–04	23.19	12–Jan–04	8.13
−6	17–Oct–03	32.08	3–Feb–04	23.26	9–Jan–04	8.15
−7	16–Oct–03	32.31	2–Feb–04	23.80	8–Jan–04	8.24
−8	15–Oct–03	31.99	30–Jan–04	24.00	7–Jan–04	8.21
−9	14–Oct–03	32.10	29–Jan–04	24.45	6–Jan–04	8.29
−10	13–Oct–03	31.87	28–Jan–04	23.67	5–Jan–04	8.03

TABLE 12.5 Target Company Stock Returns

	PRICES			RETURNS		
Date	**FBF**	**DIS**	**AWE**	**FBF**	**DIS**	**AWE**
10	39.69	26.73	11.08	−0.004	0.016	−0.01
9	39.84	26.30	11.16	−0.010	0.013	0.01
8	40.26	25.96	11.05	0.009	−0.03	0.002
7	39.89	26.75	11.03	0.007	0.008	0
6	39.60	26.55	11.02	−0.007	−0.02	−0.01
5	39.86	27.00	11.17	0.003	0.011	0.022
4	39.73	26.71	10.93	0.007	−0.01	0.03
3	39.45	26.90	10.61	0.014	−0	0.005
2	38.90	26.92	10.56	0.019	−0.04	−0.04
1	38.17	28.00	10.99	−0.010	0.014	0.058
0	38.56	27.60	10.39	0.233	0.146	0.04
−1	31.28	24.08	9.99	−0.008	0.013	0.018
−2	31.54	23.77	9.81	0.002	0.018	−0.02
−3	31.47	23.35	9.99	−0.013	0.006	0.168
−4	31.88	23.20	8.55	−0.009	0	0.052
−5	32.17	23.19	8.13	0.003	0	0
−6	32.08	23.26	8.15	−0.007	−0.02	−0.01
−7	32.31	23.80	8.24	0.010	−0.01	0.004
−8	31.99	24.00	8.21	−0.003	−0.02	−0.01
−9	32.10	24.45	8.29	0.007	0.033	0.032
−10	31.87	23.67	8.03	N/A	N/A	N/A

Note: Table calculations will reflect rounding errors.

the text Appendix A.1.a, standard deviations for each of the average daily residuals are computed along with normal deviates or test statistics ($[\varepsilon_t - 0] \div \sigma_{\varepsilon i}$) as in Table 12.8.

Our test for each daily average residual (AR_t) is structured more formally as follows:

$$H_0: AR_t \leq 0 \quad H_A: AR_t > 0$$

We shall assume the residuals follow a t-distribution and we will perform a one-tailed test with a 95% level of significance. Given a sample size of three firms such that we work with $1 = 3 - 2$ degrees of freedom, the critical value for each test will be 6.314. Based on our computations above, we find that none of the residual *t*-statistics (normal deviates or

TABLE 12.6 Returns on the S&P 500

Calendar Date	Adj. Close	Return	DIS	AWE	Calendar Date	Adj. Close	Return	FBF	AWE	Calendar Date	Adj. Close	Return	FBF
26–Feb–04	1144.91	0.00108	10		27–Jan–04	1144.05	−0.0098		5	3–Nov–03	1059.02	0.00791	5
25–Feb–04	1143.67	0.00402	9		26–Jan–04	1155.37	0.01211		4	31–Oct–03	1050.71	0.0036	4
24–Feb–04	1139.09	−0.0017	8		23–Jan–04	1141.55	−0.0021		3	30–Oct–03	1046.94	−0.0011	3
23–Feb–04	1140.99	−0.0027	7		22–Jan–04	1143.94	−0.0032		2	29–Oct–03	1048.11	0.00126	2
20–Feb–04	1144.11	−0.0026	6		21–Jan–04	1147.62	0.00777		1	28–Oct–03	1046.79	0.01519	1
19–Feb–04	1147.06	−0.0041	5		20–Jan–04	1138.77	−0.0009		0	27–Oct–03	1031.13	0.00216	0
18–Feb–04	1151.82	−0.0045	4		16–Jan–04	1139.83	0.00687		−1	24–Oct–03	1028.91	−0.0047	−1
17–Feb–04	1156.99	0.00976	3		15–Jan–04	1132.05	0.00135		−2	23–Oct–03	1033.77	0.00331	−2
13–Feb–04	1145.81	−0.0055	2		14–Jan–04	1130.52	0.00829		−3	22–Oct–03	1030.36	−0.015	−3
12–Feb–04	1152.11	−0.0049	1		13–Jan–04	1121.22	−0.0053		−4	21–Oct–03	1046.03	0.00129	−4
11–Feb–04	1,157.76	0.01067	0		12–Jan–04	1127.23	0.00479		−5	20–Oct–03	1044.68	0.00516	−5
10–Feb–04	1145.54	0.00503	−1		9–Jan–04	1121.86	−0.0089		−6	17–Oct–03	1039.32	−0.0102	−6
9–Feb–04	1139.81	−0.0026	−2		8–Jan–04	1131.92	0.00496		−7	16–Oct–03	1050.07	0.00316	−7
6–Feb–04	1142.76	0.01256	−3		7–Jan–04	1126.33	0.00237		−8	15–Oct–03	1046.76	−0.0026	−8
5–Feb–04	1128.59	0.00184	−4		6–Jan–04	1123.67	0.00129		−9	14–Oct–03	1049.48	0.00395	−9
4–Feb–04	1126.52	−0.0084	−5		5–Jan–04	1122.22	0.0124		−10	13–Oct–03	1045.35	N/A	−10
3–Feb–04	1136.03	0.00068	−6	10	10–Nov–03	1047.11	−0.0058	10					
2–Feb–04	1135.26	0.00365	−7	9	7–Nov–03	1053.21	−0.0046	9					
30–Jan–04	1131.13	−0.0026	−8	8	6–Nov–03	1058.05	0.00593	8					
29–Jan–04	1134.11	0.00499	−9	7	5–Nov–03	1051.81	−0.0014	7					
28–Jan–04	1128.48	−0.0136	−10	6	4–Nov–03	1053.25	−0.0054	6					

TABLE 12.7 Target Firm Stock Residuals

	Abnormal Returns			Average
Date	**FBF**	**DIS**	**AWE**	**Residuals (ARs)**
10	0.002	0.0153	−0.0078	0.003149
9	−0.006	0.0091	0.0063	0.003174
8	0.0033	−0.028	0.00444	−0.006695
7	0.0087	0.0103	−0.0041	0.004956
6	−0.001	−0.014	0.00018	−0.004996
5	−0.005	0.015	0.03176	0.014036
4	0.0035	−0.003	0.01805	0.006319
3	0.0153	−0.011	0.00682	0.00386
2	0.0179	−0.033	−0.0359	−0.017053
1	−0.025	0.0194	0.04998	0.014683
0	0.2306	0.1355	0.04097	0.135687
−1	−0.004	0.008	0.01148	0.005316
−2	−0.001	0.0206	−0.0194	0
−3	0.0021	−0.006	0.16013	0.052052
−4	−0.01	−0.001	0.05699	0.015093
−5	−0.002	0.0054	−0.0072	−0.00141
−6	0.0031	−0.023	−0.002	−0.007428
−7	0.0068	−0.012	−0.0013	−0.002151
−8	0	−0.016	−0.012	−0.009543
−9	0.0033	0.028	0.03109	0.020772
−10	N/A	N/A	N/A	N/A

test statistics) exceed 6.314. Thus, we may not conclude with a 95% level of confidence that any residual differs from zero. Based on the confines of the test that we established here, we may not conclude that markets appear inefficient with respect to merger announcements. Perhaps, in part due to our small sample with such a small number of degrees of freedom, we cannot conclude that merger announcements have any effect on security returns. Note that this example was structured so as to facilitate computations; it is unlikely that a realistic test would be structured with a sample set of only three firms.

The tests performed above were concerned with whether merger announcements significantly affected stock prices in any given date around the time of the announcement. We found no significant effect for any single-day returns. In some other instances, we might

TABLE 12.8 Target Firm Average Residuals and Standard Deviations

Day	Average Residuals (ARs)	σ Residuals	Normal Deviate
10	0.003149	0.0116	0.2715
9	0.003174	0.00794	0.39957
8	−0.006695	0.01834	−0.3649
7	0.004956	0.00787	0.63007
6	−0.004996	0.0079	−0.6321
5	0.014036	0.01821	0.77059
4	0.006319	0.01061	0.59555
3	0.00386	0.01313	0.29393
2	−0.017053	0.03027	−0.5633
1	0.014683	0.03786	0.38784
0	0.135687	0.0948	1.43123
−1	0.005316	0.00786	0.67597
−2	0	0.01999	0.00187
−3	0.052052	0.09369	0.55561
−4	0.015093	0.03656	0.41285
−5	−0.00141	0.00635	−0.2219
−6	−0.007428	0.01404	−0.5289
−7	−0.002151	0.00944	−0.2278
−8	−0.009543	0.00777	−1.2279
−9	0.020772	0.01524	1.36292
−10	N/A	N/A	N/A

find that while no effect is found on the residual for any particular date, the effect might be realized over a period of days. This might be expected if market reactions are slow, that is, if the market is somewhat inefficient. Perhaps, we may even wish to broaden our test to determine whether some of the effect might be realized over a period of time before the date of the announcement. We can compute cumulative average residuals to determine cumulative effects over time:

$$CAR_t = \sum_{i}^{t} AR_i$$

Cumulative average residuals are computed in Table 12.9 from average residuals taken from Table 12.8. Cumulative average residuals can also be computed by summing individual firm residuals and dividing by the number of firms in the sample as in Table 12.10.

TABLE 12.9 Target Firm Cumulative Average Residuals

Day	Average Residuals (ARs)	CAR
10	0.003149	0.22986
9	0.003174	0.22671
8	−0.006695	0.22354
7	0.004956	0.23023
6	−0.004996	0.22527
5	0.014036	0.23027
4	0.006319	0.21623
3	0.00386	0.20991
2	−0.017053	0.20606
1	0.014683	0.22311
0	0.135687	0.20843
−1	0.005316	0.07274
−2	0.000037	0.06742
−3	0.052052	0.06739
−4	0.015093	0.01533
−5	−0.00141	0.00024
−6	−0.007428	0.00165
−7	−0.002151	0.00908
−8	−0.009543	0.01123
−9	0.020772	0.02077
−10	N/A	N/A

Do any of the CARs in Table 12.9 or Table 12.10 exceed zero? A quick glance reveals that all do. Do any CARs exceed zero at a statistically significant level? We need a little more analysis to examine this second question. We begin to test for the statistical significance of cumulative average residuals by computing standard deviations of the cumulative residuals of the firms for each day and computing normal deviates. For example, the sample standard deviation of cumulative residuals for day −5 is computed based on the following:

$$\sigma = \sqrt{\frac{(0.010039-0.00024)^2 + (-0.0178-0.00024)^2 + (0.008485-0.00024)^2}{3-1}} = 0.015645$$

TABLE 12.10 Target Firm Cumulative Average Residuals

Day	CR FBF	CR DIS	CR AWE	CAR	CRσ	Normal Deviate
10	0.241609	0.1196	0.328365	0.229858	0.104878	2.191678
9	0.239582	0.104334	0.336212	0.226709	0.116474	1.946444
8	0.24544	0.095258	0.329908	0.223535	0.118849	1.880836
7	0.242097	0.123125	0.325467	0.23023	0.101692	2.263999
6	0.233407	0.112865	0.329549	0.225274	0.108571	2.074906
5	0.234481	0.12696	0.329368	0.23027	0.10127	2.273827
4	0.239118	0.11197	0.297613	0.216234	0.094913	2.278221
3	0.235621	0.114565	0.279559	0.209915	0.085448	2.456643
2	0.220366	0.125065	0.272735	0.206055	0.074868	2.752261
1	0.202502	0.158168	0.308654	0.223108	0.07733	2.885135
0	0.227803	0.138796	0.258678	0.208426	0.062246	3.348411
−1	−0.00278	0.003284	0.217708	0.072739	0.125584	0.579205
−2	0.000767	−0.00473	0.206232	0.067423	0.120244	0.560715
−3	0.001852	−0.0253	0.225603	0.067385	0.137692	0.489392
−4	−0.00027	−0.01921	0.065477	0.015333	0.044446	0.34498
−5	0.010039	−0.0178	0.008485	0.00024	0.015645	0.01535
−6	0.012391	−0.02317	0.015725	0.00165	0.021555	0.076564
−7	0.009272	0.000202	0.01776	0.009078	0.008781	1.033889
−8	0.002431	0.012187	0.019069	0.011229	0.00836	1.343124
−9	0.003266	0.027964	0.031086	0.020772	0.015241	1.362916
−10	N/A	N/A	N/A	N/A	N/A	N/A

These one-day standard deviations measure the spread or variability of residuals for that day. In Table 12.9, the normal deviate for a given date is simply the cumulative average residual for that date divided by the standard deviation applicable to that date. Daily standard deviations of cumulative residuals along with their normal deviates are given in Table 12.10. Larger normal deviates are consistent with larger positive and statistically significant stock price reactions.

The largest normal deviate is for day zero, the merger announcement date. In a realistic scenario, this would not be a surprise if prices rapidly adjusted to new information. However, we need a benchmark for statistical significance for these normal deviates. Because our sample set numbers only 3, we have $3 - 2 = 1$ degrees of freedom. For a one-tail test with 95% certainty, we find from a *t*-table that our benchmark for statistical significance is

6.314. Note that none of our normal deviates exceed this critical value of 6.314. Thus, if our hypotheses concerning each date t in our testing period were given as follows:

$$H_0: CAR_t \leq 0 \quad H_A: CAR_t > 0$$

we would not be able to reject the null hypothesis that $CAR_t \leq 0$ with 95% confidence for any date in our event period. However, we note that the sample set for our illustration was particularly small.

12.6 STRONG FORM EFFICIENCY AND INSIDER TRADING

Strong form market efficiency tests are concerned with whether any information, publicly available or private, can be used to generate abnormal returns. We generally take it for granted that insiders are capable of generating higher than normal returns on their illegal insider transactions. There have been many well-publicized cases of insiders generating abnormal returns—and being prosecuted for their trading. There is even some evidence that insiders are able to generate abnormal returns on apparently legal transactions that are duly registered with the SEC (see Jaffee, 1974). Jaffee examined SEC insider transaction filings and determined that stock performance relative to market performance after months when insider purchases exceed insider sales. When insiders sell, shares that they sold are outperformed by the market.

Why do insiders appear to outperform the market on their duly registered insider transactions? Are insiders trading on the basis of their private information or do they actually have superior trading ability? A study by Givoly and Palmon (1985) suggests that transactions generating these superior returns are not related to subsequent corporate events or announcements. Givoly and Palmon examined insider transactions and then searched for subsequent announcements in the business press for that might have explained the superior returns on their investments. They found that insider superior returns were not explained by the published announcements. This may suggest that these insiders may either simply have superior investing ability or may generate higher returns for themselves on the basis of information that is not later announced.

On the other hand, perhaps insiders are trading on the basis of insider information that is not subsequently released on a specific date. Furthermore, managers are not obliged to announce most types of inside information according to any particular schedule. In addition, many insiders participate in plans to regularly buy (without liability, as per SEC Rule 10b5-1) or sell shares. Managers can obtain 10b5-1 protection for trades if they create the plan at a time when they don't have nonpublic information and they announce their transactions schedule in advance. For example, Kenneth Lay was said to have protected $100 million in his own wealth by selling shares of Enron stock through a 10b5-1 plan. In addition, insiders always have the right to abstain from trading on the basis of inside information. Thus, it is not illegal to *not buy* shares on the basis of adverse inside information.[3]

[3]The decision to retain stock is not be a violation of SEC Rule 10b-5. See Condus v. Howard Savings Bank, 781F. Supp. 1052 (D.N.J. 1992). However, failure to trade can be a violation of U.K. insider trading regulations.

How would investigators determine whether one declined to trade solely on the basis of inside information? Regardless, Jagolinzer (2009) found that insider trading within the 10b5-1 plans outperforms the market by 5.6% over six-month periods.

Consider securities trading by underage investors, which, based on NASDAQ OMX Helsinki exchange data, has been demonstrated to produce higher than normal returns (Berkman, Koch, & Westerholm, 2014). While all of us raise children who are smarter than average, Berkman et al. argue that such trades are initiated by guardians, that they are more likely to be profitable than trades initiated by other adult traders and that such trades are particularly profitable, as high as 12% per day when executed before major earnings and takeover announcements or large share price changes. Guardians executing these trades, Berkman et al. suggest, are "likely to be above-average investors," and have "superior cognitive skills or comparative advantages in obtaining value-relevant information." In addition, guardians' trades on children's accounts outperform trades on their own accounts, and might camouflage illegal trading or reduce the information-related price impact of their trade executions.

Children aren't the only investors that might benefit systematically from non-public information. Schweizer (2011) details numerous cases of members of the U.S. House of Representatives profiting from trades while holding non-public information concerning issues pending before Congress. For example, former House Speaker John Boehner reportedly purchased stock in health insurance companies just prior to the failure of the "Obamacare public option," which served to increase health insurance stock prices. Ziobrowski, Boyd, Cheng, and Ziobrowski (2004, 2011) found that a U.S. stock portfolio imitating transactions of members of the U.S. House of Representatives and Senate during the mid- and late-1990s produced annual abnormal returns of approximately 10%.[4] As a result of these and other similar findings, Congress passed its 2012 STOCK Act (Stop Trading on Congressional Knowledge), which prohibits Congressmen, their families and their staffs from trading on the basis of non-public privileged information obtained through their Congressional service.

Even U.S. securities regulators seem able to produce higher than normal returns on apparent non-public information. For example, Rajgopal and White (2017) found that a hypothetical U.S. stock portfolio imitating transactions of S.E.C. employees produce positive annual abnormal returns of approximately 8.5%. These returns are driven entirely by avoiding losses on the sell-side before significant price declines. Perhaps these abnormal returns are due to good fortune or a statistical artifact. Perhaps they are due to superior stock-picking skills of S.E.C. employees. However, S.E.C. employees are often instructed to sell from their portfolios shares in firms when they are assigned to secret investigations. Oddly, the S.E.C. might require that their employees sell stock on the basis of non-public adverse information, aiding employees in their efforts to avoid trading losses.

[4]On the other hand, Eggers and Hainsmueller (2013) found that members of Congress and the Senate were poor investors, underperforming the market by about 2%−3% per year. However, Ziobrowski's Senate results appear to have been driven by a minority of senators, with only four senators accounting for nearly half of all sampled stock transactions. Perhaps the four more aggressive-trading legislators produced better insider-trading results.

12.7 ANOMALOUS EFFICIENCY AND PREDICTION MARKETS

Sometimes markets seem so efficient that prices can be used to predict nonfinancial events. For example, French and Roll (1986) found that prices of orange juice concentrate futures can be used to forecast temperature conditions in central Florida better than the National Weather Service forecasts. In other instances, the market incorporates relevant information so efficiently that it is difficult to understand how the process worked.

The *Challenger* Space Shuttle Disaster

On January 28, 1986, at 11:38 AM Eastern Standard Time, the space shuttle *Challenger* was launched in Florida and exploded 74 seconds later ten miles above ground on national television. The stock market reacted within minutes of the event, with investors dumping shares of the four major contractors contributing to building and launching the *Challenger*. The four primary contractors, Rockwell International, builder of the shuttle and its main engines, Lockheed, manager of the ground support, Martin Marietta, manufacturer of the vessel's external fuel tank, and Morton Thiokol, builder of the solid-fuel booster rocket. Less than a half-hour after the disaster, Rockwell's stock price had declined 6%, Lockheed 5%, Martin Marietta 3%, and Morton Thiokol had stopped trading because of the flood of sell orders. By the end of trading for the day, the first three companies' share prices closed down 3% from their open prices, representing a slight recovery from their initial reactions. However, Morton Thiokol stock resumed trading and continued to decline, finishing the day almost 12% down from its open price. These reactions suggested that the market believed that Morton Thiokol would suffer the greatest losses from the disaster, despite the fact that no reports surfaced in the public media identifying Morton Thiokol as the cause of the disaster. Even news reports of rumors in the media failed to single out the firm as the cause of the disaster.

However, many months after the disaster, Richard Feynman, the charismatic and brilliant physicist, in dramatic testimony to a congressional hearing on the explosion, dropped O-rings into ice water, causing them to become brittle and easily break, thereby demonstrating that they were the cause of the explosion. Morton had used the O-rings in its construction of the booster rockets, which failed in the uncharacteristically cold weather that morning and leaked explosive fumes when the launch temperatures were less than could be tolerated by the O-rings. Yet, there were no announcements of such failures on the dates of the disaster or even within weeks of the explosion. Nonetheless, the market had reacted within minutes of the disaster as though Morton Thiokol would be held responsible.

In their study of this event, Maloney and Mulherin (2003) found no evidence that Morton Thiokol corporate officers and other insiders sold shares on the date of the disaster. How is it that no individual seemed to know that Morton Thiokol would ultimately be held responsible for the disaster, yet the market would react as though everyone knew? It may be useful to note that Morton Thiokol engineers were aware of the potential for failure of the O-rings in cold weather, but were overruled by company managers concerning their use. But, again, Maloney and Mulherin found no evidence of unusual insider trading activity. This scenario suggests that the information marketplace does quickly sift through and identify relevant information used for the valuation of company shares.

Prediction Markets

Price discovery is one of the most important functions of trading, particularly in more transparent markets such as the NYSE. This price discovery is of vital importance to investors and business decision making. What is the essence of trading that produces this information? Several of the models that we have discussed in this chapter allow for market frameworks where informed traders communicate information (perhaps reluctantly) through their trading activities. Might gambling serve the same essential purpose as trading in the price discovery and information production processes? Similarly, what might gambling be able to teach us about information dissemination in financial markets? Bear in mind that prediction markets encourage and reward participation by more knowledgeable traders and punish noise traders.

Consider the recent presidential elections, where an increasing number of online betting markets offered tradable securities on election outcomes.[5] Among the most visible of these markets have been PredictIt and the Iowa Electronic Markets (IEM) at the Tippie School of Business, University of Iowa. Both have created "futures" contracts that can be purchased and sold by "investors," whose values fluctuated as perceived election chances of various candidates and referendums varied. Thus, a contract that pays $1 if a given candidate is elected might sell for a price less than $1. For example, if a contract sells for $0.50, one might roughly infer that the market believes that the candidate has a 50% chance (ignoring discount factors and other complications) of getting elected. In effect, futures contract prices might have been interpreted as the market's assessment of the probability that a given candidate would be elected. Hence, such a site might serve as a *prediction market*. Prediction markets are markets created for the purpose of making predictions or benefiting from correct predictions. The Iowa Electronic Markets, PredictIt and Intrade (defunct as of early 2013 following a CFTC lawsuit) have been more successful at predicting election outcomes than many opinion polls. Consider, for example, the report by *Time Magazine* (Saporito, 2005) of Intrade's contract making payment contingent on the capture of Saddam Hussein suddenly moving two days before he was captured on December 13, 2003. Other prediction markets such as Betfair, Cantor Fitzgerald's Hollywood Stock Exchange, Centrebet (now, William Hill), Goldman Sach's, and Deutsche Bank's Economic Derivatives and TradeSports (which became InTrade) have served as useful prediction devices in other arenas, and have been very useful in making nonfinancial predictions. On the other hand, prices on some particularly high-volume prediction markets seemed inconsistent with actual outcomes. The 2016 Brexit (Britain's exit from the European Union) and the 2016 U.S. presidential election, in the Republican primaries and the general election, both ended in results contrary to *apparent* market predictions on PredictIt and IEM. At a minimum, we need to remember that these were not the only prediction markets operating at the time, markets produce implied-probabilities rather than actual forecasts and that both actual voting contests were quite close. A number of research papers have discussed

[5]Organized betting markets in political elections preceded the Internet, even on Wall Street. Although such gambling has recently been either illegal or prohibited by NYSE and AMEX rules (but not always, as we will discuss shortly), well-organized "underground" markets have existed for over a hundred years on the major U.S. exchanges.

use of prediction markets for a variety of types of predictions (e.g., Wolfers and Zitzewitz, 2004, 2007).[6]

Security markets are excellent aggregators of information. Security prices have been used for many years to estimate a variety of types of probability distributions. For example, currency traders have for many years used futures prices to estimate future currency exchange rates. Similarly, commodity traders have used commodity futures prices to predict commodity prices. Latane and Rendleman (1976) proposed used market prices of put and call options to estimate implied volatilities (return standard deviations) for underlying stocks (discussed in Chapter 7). Bodurtha and Shen (1995) and Campa and Chang (1998) used similar statistical methods to estimate implied correlations between two underlying variables such as exchange rates using derivative contracts written on each underlying currency as well as contracts written on both currencies.

Prediction markets, even with respect to political wagering, did not originate with Intrade and the Iowa Electronic Markets. The Curb Exchange (the precursor to the American Stock Exchange) operated wagering markets for presidential markets during much of the late 19th century. Major newspapers, including The New York Times regularly printed gambling odds on elections. Political wagering frequently involved large sums of money, with daily volume that often exceeded presidential campaign budgets. More recent prediction markets have been quite successful, including the North American Derivatives Exchange (Nadex), a Commodity Futures Trading Commission (CFTC)-registered futures exchange that got its start as HedgeStreet prediction market.

We will focus here on several questions. Is the information provided by markets of use to traders? Clearly, derivative markets do provide useful information for forecasting risk, futures rates, and so on. Is the information provided by markets of use to decision-making entities in business and government? Consider an example from the 1990s where CERN, the European laboratory for particle physics, needed to estimate whether the probability of discovering the Higgs boson was sufficiently high to justify extending the operation of its collider.[7] Traders at the Foresight Exchange website (http://www.ideosphere.com/) took positions on whether the Higgs boson would be discovered by 2005, setting a contract price of 0.77 as of 2001. We close this section with a few rhetorical questions: Should markets provide information aggregation services to the public? If so, at what cost to traders? Consider the following excerpt from Looney (2003)[8]:

> The Defense Advanced Research Projects Agency (DARPA) was born in the uncertain days after the Soviets launched Sputnik in 1958. Its mission was to become an engine of technological change that would bridge the gap between fundamental discoveries and their military use (Bray, 2003). Over the last five decades, the Agency has efficiently gone about its business in relative obscurity, in many cases not getting as much credit as it deserved. The Agency first developed the model for the internet as well as stealth

[6] Wolfers and Zitzewitz (2004) was the source for some of the information in this section.

[7] The boson is a subatomic particle (such as a quark or leptons) that seems to impart mass on other bosons, quarks, and leptons. While there was strong theoretical and some empirical evidence for the existence for the Higgs boson as of June 2010, its existence had not been empirically confirmed at that time.

[8] Excerpted from http://www.au.af.mil/au/awc/awcgate/nps/pam/si_pam.pdf on April 12, 2012.

technology. More recently, DARPA innovations have spanned a wide array of technologies. To name a couple: computers that correct a user's mistakes or fix themselves when they malfunction and new stimulants to keep soldiers awake and alert for seven consecutive days.

Because DARPA is mandated to take on risky projects, failures have occurred. For the most part, however, the Agency's low profile has protected it from inaccurate scare stories cropping up in the popular press. However, in 2003 DARPA has managed to make the front pages twice, both times with disastrous results. Earlier in the year Congress moved to scale back the agency's Terrorism Information Awareness Program (TIA). In an effort to spot patterns of terrorist activity, this program proposed the development of advanced computer systems capable of scanning commercial databases containing information on millions of Americans.

Then, in late July, the Agency backed off a plan to set up a kind of futures market (Policy Analysis Market or PAM) that would allow investors to earn profits by betting on the likelihood of such events as regime changes in the Middle East. Critics, mainly politicians and op-ed writers, attacked the futures project on the grounds that it was unethical and in bad taste to accept wagers on the fate of foreign leaders and the likelihood of terrorist attacks. The project was canceled a day after it was announced. Its head, retired Admiral John Poindexter, has resigned.

One addition to this excerpt: Poindexter's resignation followed the creation of a contract by Tradesports.com that would pay $100 if he resigned. In any case, can markets trading terrorism-based contracts aid in the prediction of terrorist strikes and dealing with the effects of such strikes? If so, should such contracts be traded? For example, might the risk associated with investment in military or private security companies be managed with contracts on specific terrorist actions? Should such contracts be traded even in the light of potential moral hazard consequences?

12.8 EPILOGUE

In his presidential address to the American Finance Association, Richard Roll (1988) discussed the ability of academics to explain financial phenomena:

The maturity of a science is often gauged by its success in predicting important phenomena. Astronomy, the oldest science, is able to predict the positions of planets and the reappearance of comets with a high degree of accuracy. . . . The immaturity of our science [finance] is illustrated by the conspicuous lack of predictive content about some of its most intensely interesting phenomena, particularly changes in asset prices. General stock price movements are notoriously unpredictable and financial economists have even developed a coherent theory (the theory of efficient markets) to explain why they should be unpredictable.

The theories of capital market efficiency are quite powerful in that they are quite simple and do explain much of the behavior that is observed in capital markets. However, much empirical evidence exists that refutes capital markets efficiency. It should be noted that much of this evidence is contradictory and often reflects investors' desire to discover money-making strategies, investment advisors' needs to sell their services, and academicians' needs for publications. Furthermore, some evidence of market inefficiencies may simply be the result of data mining; one can always find or "demonstrate" an interesting

pattern or relationship given enough data. An even greater difficulty for the opponents of capital market efficiency is that they are not able to offer a reasonably coherent and robust set of competing theories of capital markets behavior. However, a good start to making this argument might be to consider the scenario of Grossman and Stiglitz (1980) who argue that markets cannot be fully informationally efficient because no one in such markets could earn abnormal returns or would have incentives to obtain and analyze information. Hence, there would be no mechanism to create informationally efficient markets to begin with. Thus, it is reasonable to conclude based on prevailing theoretical and empirical research that the verdict is still out on the extent to which capital markets are efficient.

Additional Reading

Perhaps the single most important book for any investor or trader to read is Burton Malkiel's still-relevant classic *A Random Walk Down Wall Street* (2012). It is a nontechnical easy read and provides a rather skeptical discussion of a wide variety of trading strategies, and has been a financial best-seller since the early 1070s. Lo and MacKinlay's (1999) more technical response, *A Non-Random Walk Down Wall Street*, is a more academic collection of essays depicting the nonrandom nature of stock prices. Two books concerning technical analysis are particularly good reads. The first, Edwards, Magee, and Bassetti (2012), or an earlier edition thereof, offers a basic how-to guide for charting and technical analysis. The second, a collection of essays edited by Acar and Satchell (2002) discuss more quantitative and algorithmic approaches to technical analysis. One essay, by Lo, Mamaysky, and Wang (2000), reprinted from the *Journal of Finance*, discusses their empirical results on several technical tools and their approaches to designing algorithms for finding and testing transactions opportunities. Chapter 17 of Elton, Gruber, Brown, and Goetzman (2010) provides a nice review of the market efficiency literature and discusses the methodology of event studies. Chapters 2 and 4 of Campbell, Lo, and MacKinlay (1997) discuss the econometrics of testing stock return randomness and event studies. Wolfers and Zitzewitz (2004, 2007) offer very useful and readable introductions to prediction markets. Surowiecki's 2004 well-written book on crowd wisdom provides good insight on how divergent and largely uninformed individual opinions are aggregated into a highly efficient market. Boudoukh, Richardson, and Whitelaw (1994) offer an interesting synthesis of the literature on mean reversion and momentum trading rules.

References

Acar, E, & Satchell, S. (2002). *Advanced trading rules*. Woburn, MA: Butterworth Heinemann.

Alexander, S. (1961). Price movements in speculative markets: Trends or random walks. *Industrial Management Review*, *2*, 7–26.

Ball, R., & Brown, P. (1968). An empirical evaluation of accounting income numbers. *Journal of Accounting Research*, *6*, 159–178.

Banz, R. (1981). The relationship between return and market value of common stock. *Journal of Financial Economics*, *9*, 3–18.

Barber, B., Lehavy, R., & Truman, B. (2007). Comparing the stock recommendation performance of investment banks and independent research firms. *Journal of Financial Economics*, *85*, 490–517.

Basu, S. (1977). Investment performance of common stocks in relation to their price-earnings ratios. *Journal of Finance*, *32*, 663–682.

Berkman, H., Koch, P., & Westerholm, J. (2014). Informed trading through the accounts of children. *Journal of Finance*, *69*, 363–404.

Black, F. (1973). Yes Virginia, there is hope: Tests of the Value Line ranking system. *Financial Analysts Journal*, *29*, 10–14.

Bodurtha, J.N., & Shen, Q. (1995). *Historical and implied measures of "value at risk": The DM and yen case*. Working paper. Ann Arbor: University of Michigan.

Boudoukh, J., Richardson, M., & Whitelaw, R. (1994). A tale of three schools: Insights on autocorrelations of short-horizon stock returns. *Review of Financial Studies*, *7*, 539–573.
Branch, B. (1977). A tax loss selling rule. *Journal of Business*, *50*, 198–207.
Bray, H. (2003, August 3). *In defense of DARPA programs*. Boston Globe.
Brock, W., Lakonishok, J., & LeBaron, B. (1992). Simple technical trading rules and the stochastic properties of stock returns. *Journal of Finance*, *47*, 1731–1764.
Brown, S., & Warner, J. (1980). Measuring security price performance. *Journal of Financial Economics*, *8*, 205–258.
Brown, S., & Warner, J. (1985). Using daily stock returns: The case of event studies. *Journal of Financial Economics*, *14*, 3–31.
Brown, W. O., & Sauer, R. D. (1993). Fundamentals or noise? Evidence from the professional basketball betting market. *Journal of Finance*, *48*, 1193–1209.
Campa, J. M., & Chang, P. H. K. (1998). The forecasting ability of correlations implied in foreign exchange options. *Journal of International Money and Finance*, *17*, 855–880.
Campbell, J. Y., Lo, A. W., & MacKinlay, A. C. (1997). *The econometrics of financial markets*. Princeton, NJ: Princeton University Press.
Chan, K. C., Chen, N. F., & Hsieh, D. (1985). An exploratory investigation of the firm size effect. *Journal of Financial Economics*, *14*, 451–471.
Chang, E. C., & Pinegar, J. M. (1986). Return seasonality and tax-loss selling in the market for long term government and corporate bonds. *Journal of Financial Economics*, *17*, 391–415.
Cowles, A., 3rd. (1933). Can stock market forecasters forecast? *Econometrica*, *1*, 309–324.
Cox, G. V. (1930). *An appraisal of American business forecasts*. Chicago: University of Chicago Press.
Davies, P. L., & Canes, M. (1978). Stock prices and the publication of second hand information. *Journal of Business*, *51*, 43–56.
DeBondt, W. F. M., & Thaler, R. (1985). Does the market overreact? *Journal of Finance*, *40*, 793–805.
Dyl, E. (1977). Capital gains taxation and year-end stock market behavior. *Journal of Finance*, *32*, 165–175.
Edwards, R. D., Magee, J., & Bassetti, W. H. C. (2012). *Technical analysis of stock trends*. Boca Raton, FL: CRC Press.
Eggers, A. C., & Hainmueller, J. (2013). Capitol losses: The mediocre performance of congressional stock portfolios, 2004–2008. *Journal of Politics*, *75*(2), 535–551.
Elton, E., Gruber, M., Brown, S., & Goetzman, W. (2010). *Modern portfolio theory and investment analysis* (8th ed.). New York: Wiley.
Fama, E. (1965). The behavior of stock market prices. *Journal of Business*, *38*, 34–105.
Fama, E., & Blume, M. (1966). Filter rules and stock market trading profits. *Journal of Business*, *39*, 226–241.
Fama, E., & French, K. (1992). The cross-section of expected stock returns. *Journal of Finance*, *47*, 427–465.
Fama, E., & MacBeth, J. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, *71*, 607–636.
Fama, E., Fisher, L., Jensen, M. C., & Roll, R. (1969). The adjustment of stock prices to new information. *International Economic Review*, *10*, 1–21.
Firth, M. (1976). The impact of earnings announcements on the share price behaviour of similar type firms. *Economics Journal*, *86*, 296–306.
French, K. (1980). Stock returns and the weekend effect. *Journal of Financial Economics*, *8*, 55–69.
French, K., & Roll, R. (1986). The arrival of information and the reaction of traders. *Journal of Financial Economics*, *13*, 547–559.
Gandar, J., Zuber, R., O'Brien, T., & Russo, B. (1988). Testing rationality in the point spread betting market. *Journal of Finance*, *43*, 995–1008.
Gao, M., & Huang, J. (2016). Capitalizing on Capitol Hill: Informed trading by hedge fund managers. *Journal of Financial Economics*, *121*, 521–545.
Givoly, D., & Ovadia, A. (1983). Year-end induced sales and stock market seasonality. *Journal of Finance*, *38*, 171–185.
Givoly, D., & Palmon, O. (1985). Insider trading and the exploitation of inside information: Some empirical evidence. *Journal of Business*, *58*, 69–87.
Golec, J., & Tamarkin, M. (1991). The degree of inefficiency in the football betting markets. *Journal of Financial Economics*, *30*, 311–323.

Granger, C. W. (1968). Some aspects of the random walk model of stock prices. *International Economic Review*, *9*, 253–257.
Granger, C.W., & Morgenstern, O. (1972). *Predictability of stock market prices*. Boston: Heath.
Gray, P., & Gray, S. F. (1997). Testing market inefficiency: Evidence from the NFL sports betting market. *Journal of Finance*, *52*, 1725–1737.
Green, T. C. (2006). The value of client access to analyst recommendations. *Journal of Financial and Quantitative Analysis*, *41*, 1–24.
Griffiths, M., & White, R. (1993). Tax-induced trading and the turn of the year anomaly: An intraday study. *Journal of Finance*, *48*, 575–598.
Grossman, S., & Stiglitz, J. (1980). On the impossibility of informationally efficient markets. *American Economic Review*, *70*, 393–408.
Holloway, C. (1981). A note on testing an autoregressive investment strategy using value line ranks. *Journal of Finance*, *36*, 711–719.
Huberman, G., & Kandel, S. (1987). Market efficiency and value line record. *Journal of Business*, *63*(2), 187–216.
Ibbotson, R., & Jaffee, J. (1975). Hot issue markets. *Journal of Finance*, *30*(2), 1027–1042.
Jaffee, J. (1974). Special information and insider trading. *Journal of Business*, *47*, 410–428.
Jaffee, J., Keim, D. H., & Westerfield, R. (1989). Earnings yields, market values, and stock returns. *Journal of Finance*, *44*, 135–148.
Jagolinzer, A. D. (2009). SEC rule 10b5-1 and insiders' strategic trade. *Management Science*, *55*, 224–239.
Jegadeesh, N. (1991). Seasonality in stock price mean: Evidence from the U.S., the U.K. *Journal of Finance*, *46*, 1427–1444.
Jegadeesh, N., & Titman, S. (1993). The returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance*, *48*, 65–91.
Jennergren, P., & Korsvold, P. (1975). The non-random character of Norwegian and Swedish stock market prices. In E. J. Elton, & M. J. Gruber (Eds.), *International capital markets* (pp. 37–54). Amsterdam: North-Holland.
Jensen, M., & Bennington, G. (1970). Random walks and technical theories: Some additional evidence. *Journal of Finance*, *25*, 469–482.
Kaplan, S. N., & Ruback, R. S. (1995). The valuation of cash flow forecasts: An empirical analysis. *The Journal of Finance*, *50*(4), 1059–1093.
Keim, D. B. (1983). Size-related anomalies and stock return seasonality: Further empirical evidence. *Journal of Financial Economics*, *12*, 13–32.
Keim, D. B. (1989). Trading patterns, bid-ask spreads, and estimated security returns: The case of common stock at calendar turning points. *Journal of Financial Economics*, *25*, 75–97.
Kendall, M. (1953). The analysis of economic time series. Part I: Prices. *Journal of the Royal Statistical Society*, *96*, 11–25.
Kihn, J. (1996). The financial performance of low-grade municipal bond funds. *Financial Management*, *25*, 52–73.
Koudijs, P. (2015). Those who know most: Insider trading in 18th century Amsterdam. *The Journal of Political Economy*, *123*, 1356–1409.
Koudijs, P. (2016). The boats that did not sail: Asset price volatility in a natural experiment. *Journal of Finance*, *71*(3), 1185–1226.
Latane, H., & Rendleman, R. (1976). Standard deviations of stock price ratios implied in option prices. *Journal of Finance*, *31*, 369–381.
Levy, R. (1967). Relative strength as a criterion for investment selection. *Journal of Finance*, *22*, 595–610.
Lo, A. W., & MacKinlay, A. C. (1988). Stock market prices do not follow random walks: Evidence from a simple specification test. *Review of Financial Studies*, *1*, 41–46.
Lo, A. W., & MacKinlay, A. C. (1999). *A non-random walk down wall street*. Princeton, NJ: Princeton University Press.
Lo, A. W., Mamaysky, H., & Wang, J. (2000). Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation. *Journal of Finance*, *55*, 1705–1764.
Looney, R. (2003). DARPA's policy analysis market for intelligence: Outside the box or off the wall? *Strategic Insights*, 2. Available at <http://www.au.af.mil/au/awc/awcgate/nps/pam/si_pam.pdf> Accessed 12.04.12.

Malkiel, B. (2012). *A random walk down wall street*. New York: W.W. Norton and Company.
Maloney, M., & Mulherin, J. H. (2003). The complexity of price discovery in an efficient market: The stock market reaction to the Challenger crash. *Journal of Corporate Finance*, *9*, 453–479.
McDonald, J., & Fisher, A. K. (1972). New issue stock price behavior. *Journal of Finance*, *27*, 97–102.
Michaely, R., & Womack, K. (1999). Conflict of interest and the credibility of underwriter recommendations. *Review of Financial Studies*, *12*, 653–686.
Mullins, B., & Scannell, K. (2006). Hedge funds hire lobbyists to gather tips in Washington, Wall Street. *Journal*, *8*, A1.
Nayar, N., Singh, A. K., & Yu, W. (2011). *Unraveling a puzzle: The case of Value Line timeliness rank upgrades. Financial Markets and Portfolio Management*, *25*(4), 379–409.
Neiderhoffer, V., & Osborne, M. F. M. (1966). Market making and reversal on the stock exchange. *Journal of the American Statistical Association*, *61*, 897–916.
Rajgopal, S., & White, R. M. (2017). Stock trades of Securities and Exchange Commission employees. *Journal of Law and Economics*, *60*(3), 441–477.
Reinganum, M. (1983). The anomalous stock market behavior of small firms in January: Empirical tests for tax-loss selling effects. *Journal of Financial Economics*, *12*, 89–104.
Richards, A. J. (1997). Winner-loser reversals in national stock market indices: Can they be explained? *Journal of Finance*, *52*, 2129–2144.
Richardson, M., & Stock, J. H. (1988). Inferences drawn from statistics based on multiyear asset returns. *Journal of Financial Economics*, *25*, 323–348.
Ritter, J. (1991). The long-run performance of initial public offerings. *Journal of Finance*, *46*, 3–28.
Ritter, J. R. (1984). The "hot issue" market of 1980. *Journal of Business*, *57*, 215–240.
Roll, R. (1981). A possible explanation of the small firm effect. *Journal of Finance*, *36*, 879–888.
Roll, R. (1983). Vas ist das? The turn-of-the-year effect and the return premia of small firms. *Journal of Portfolio Management*, *9*, 18–28.
Roll, R. (1988). R^2. *Journal of Finance*, *43*, 541–566.
Rozeff, M. S., & Kinney, W. R., Jr. (1976). Capital markets seasonality: The case of stock returns. *Journal of Financial Economics*, *3*, 379–402.
Samuelson, P. (1965). Proof that properly anticipated prices fluctuate randomly. *Industrial Management Review*, *6*, 41–49.
Saporito, B. (2005). Place your bets! *Time Magazine*, October 24, 76.
Sauer, R., Brajer, V., Ferris, S. P., & Marr, M. W. (1988). Hold your bets: Another look at the efficiency of the gambling market for NFL games. *Journal of Political Economy*, *96*, 206–213.
Schweizer, P. (2011). *Throw them all out: How politicians and their friends get rich off insider stock tips, land deals, and cronyism that would send the rest of us to prison*. New York: Houghton Mifflin Harcourt.
Schwert, W. G. (1983). Size and stock returns and other empirical regularities. *Journal of Financial Economics*, *12*(1), 3–12.
Steadman, H. J., & Cocozza, J. (1978). Psychiatry, dangerousness and the repetitively violent offender. *Journal of Criminal Law and Criminology*, *69*(2), 226–231.
Shiller, R. J. (1981). Do stock prices move too much to be justified by subsequent changes in dividends? *American Economic Review*, *71*, 421–436.
Sullivan, R., Timmerman, A., & White, H. (1997). Data-snooping, technical trading rule performance, and the bootstrap. *Journal of Finance*, *54*, 1647–1691.
Surowiecki, J. (2004). *The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economics, societies, and nations*. New York: Doubleday.
Thaler, R. H., & Ziemba, W. T. (1988). Anomalies: Parimutuel betting markets: Racetracks and lotteries. *Journal of Economic Perspectives*, *2*, 161–174.
Wolfers, J., & Zitzewitz, E. (2004). Prediction markets. *Journal of Economic Perspectives*, *18*, 107–126.
Wolfers, J., & Zitzewitz, E. (2007). *Interpreting prediction market prices as probabilities*. Working paper, University of Pennsylvania. Available at <http://bpp.wharton.upenn.edu/jwolfers/Papers/InterpretingPredictionMarketPrices.pdf>.
Womack, K. L. (1996). Do brokerage analysts' recommendations have investment value? *Journal of Finance*, *51*, 137–167.

Woodland, L. M., & Woodland, B. M. (1994). Market efficiency and the favorite-longshot bias: The baseball betting market. *Journal of Finance*, *49*, 269–279.

Working, H. (1934). A random difference series for use in the analysis of time series. *Journal of the American Statistical Association*, *29*, 11–24.

Ziobrowski, A. J., Boyd, J. W., Cheng, P., & Ziobrowski, B. J. (2004). Abnormal returns from the common stock investments of the US Senate. *Journal of Financial and Quantitative Analysis*, *39*(4), 661–676.

Ziobrowski, A. J., Boyd, J. W., Cheng, P., & Ziobrowski, B. J. (2011). Abnormal returns from the common stock investments of members of the US House of Representatives. *Business and Politics*, *13*(1), 1–24.

12.9 EXERCISES

1. If price changes reflect new information, why should returns fluctuate randomly?
2. Consider the following stock price sequence:

t	P_t
1	50
2	51
3	52
4	58
5	56

   Compute three-day moving averages for days 3, 4, and 5. If one assumes that the three-day moving average price represents the true value of the stock because of the elimination of noise, should the investor buy or sell shares on days 3, 4, and 5?
3. Compute five-day moving averages for each date of the following series, assuming that 5 is the price for day 1 and 18 is the price for day 12:

5	6	5	6	8	10	9	12	15	12	16	18

4. Toy stores have very clear seasonal patterns in their revenue flows. Should toy store company share prices exhibit similar seasonality?
5. An analyst has tested the effect of spin-off announcements on share prices using event study methodology. The analyst's procedure proceeded as follows:
   a. Ten spin-offs were obtained from the most recent year.
   b. Announcement dates were standardized with 30 days of prices collected from both before and announcement dates.
   c. Returns were computed and averaged for each company in the spin-off sample. Standard deviations and normal deviates were computed for the sample for each date. Appropriate tests for statistical significance were conducted.
   d. Cumulative average residuals were computed for each date and tested for statistical significance.

Test results suggested that spin-offs were associated with 30-day returns of 8%. Why might an investor be skeptical of this test? How can the test be improved?

6. Which is worth more?
   a. A market forecasting service that correctly predicts the direction of the market portfolio with 100% consistency every month for extended periods of time and can be expected to continue to do so in the future.
   b. A market forecasting service that incorrectly predicts the direction of the market portfolio with 100% consistency every month for extended periods of time and can be expected to continue to do so in the future.
7. Suppose that we wished to conduct an event study on whether acquiring firms experience share price reactions to takeover announcements. For our event study, we will use for our sample the following three acquiring firms:
   Company X: Merger announcement date January 15, 2012
   Company Y: Merger announcement date February 15, 2012
   Company Z: Merger announcement date April 10, 2012
   Suppose we establish an 11-day testing period for returns around the event dates, the event date plus five days before and five days after. The following table provides our three acquiring firm stock prices during 12-day periods around merger announcement dates:

Company *X*		Company *Y*		Company *Z*	
Date	**Price**	**Date**	**Price**	**Date**	**Price**
1/09	50.125	2/09	20.000	4/04	60.375
1/10	50.125	2/10	20.000	4/05	60.500
1/11	50.250	2/11	20.125	4/06	60.250
1/12	50.250	2/12	20.250	4/07	60.125
1/13	50.375	2/13	20.375	4/08	60.000
1/14	50.250	2/14	20.375	4/09	60.125
1/15	52.250	2/15	21.375	4/10	60.625
1/16	52.375	2/16	21.250	4/11	60.750
1/17	52.250	2/17	21.375	4/12	60.750
1/18	52.375	2/18	21.500	4/13	60.875
1/19	52.500	2/19	21.375	4/14	60.875
1/20	52.375	2/20	21.500	4/15	60.875

   a. Compute one-day returns for each of 11 days for each of the three stocks.
   b. Suppose that we have decided to use the mean adjusted return method to compute excess or abnormal stock returns. Here, we will compute mean daily returns for each security for a period outside of our 11-day testing period. Suppose we compute average daily returns and standard deviations for each of the stocks for 180-day periods prior to the testing periods (the raw returns data are not given

here). Suppose that we have found normal or expected daily returns along with standard deviations as follows:

Stock	Normal Return	Standard Deviation
X	0.000465	0.00415
Y	0.000520	0.00637
Z	0.000082	0.00220

Compute excess returns for each stock for each of the 11 days.

c. For each of the 11 days in the analysis, compute average residuals for the three stocks. Then, for each day, compute a standard deviation of residuals for the three stocks. Finally, compute normal deviates for each of the 11 dates based on the averages and standard deviations for the three stocks.

d. Are average residuals for any of the dates statistically significant at the 95% level?

e. Compute cumulative average residuals for each of the 11 dates.

f. Compute standard deviations and normal deviates for each of the 11 dates.

g. Does there appear to be statistically significant evidence of abnormal acquiring firm returns around announcement dates?

8. The following table reflects abnormal returns for each of 10 stocks over a seven-day period about day zero, which is the standardized date of the sudden death of the CEO for each of the 10 firms. An analyst wants to determine whether the death announcement date represents a significant event and make appropriate portfolio adjustments when CEOs suddenly die in the future.

Day	Stock 1	Stock 2	Stock 3	Stock 4	Stock 5	Stock 6	Stock 7	Stock 8	Stock 9	Stock 10
−3	0.003279	−0.00814	0.008945	−0.00255	0.011395	−0.00797	0.011223	0.014037	0.020344	0.015708
−2	0.004440	−0.01997	0.017064	0.003790	0.048419	0.017966	−0.01098	−0.02468	0.023697	0.031806
−1	0.031991	0.022567	−0.01350	0.026311	0.069986	0.047213	0.012632	0.038042	0.015502	0.002453
0	0.091624	0.086352	0.088165	0.077971	0.128351	0.112391	0.069263	0.089396	0.107763	0.102206
1	0.043105	0.028967	0.027764	0.006769	−0.04901	−0.01061	0.010702	−0.00974	0.005449	0.009521
2	0.006715	0.011545	0.014355	0.013281	−0.02619	−0.03346	0.026330	0.003409	0.015373	0.006225
3	−0.01942	−0.00453	−0.00708	−0.00114	−0.00493	0.02925	−0.00783	−0.03250	0.024930	0.030536

a. What are the average residuals, their standard deviations, and normal deviates for each of the seven dates?

b. What are the cumulative average residuals, their standard deviations, and normal deviates for each of the seven dates?

c. On which days are average residuals statistically significant at the 1% level? On which days are cumulative average residuals statistically significant at the 1% level?

9. The Hollywood Stock Exchange (www.hsx.com) has described itself as "the world's leading entertainment stock market." The site has been owned by Cantor Fitzgerald, one of the world's leading trading firms. "Investors" at the HSX.trade virtual shares of celebrities and movies with virtual money called the Hollywood Dollar. For example, "investors" can trade "stock" on a new major movie release. Although the site is still considered an entertainment website, how might information concerning the price of this "stock" be of use to investors in the actual stock market? Similarly, how might information concerning the price of this "stock" be of use to managers of firms that produce and distribute films?
10. In Chapter 6, we discussed a simple dual market setting involving trading of English securities in 18th-century London markets and simultaneously in Amsterdam markets. Information would be revealed on these English securities in London markets, and this information would be delivered in Amsterdam via mail ships that departed London twice per week and required three days under the most favorable sailing conditions to reach Amsterdam. However, sometimes adverse weather would delay the arrival of these ships, and Koudijs (2015, 2016) was able to identify the impact of the delayed arrival of news, both public in newspapers and private in private letters, on Amsterdam markets. He further argued that news from England was the only information relevant to the pricing of English securities, and that the mail packets on ships were the only source of this news.
    a. When mail ship arrivals were timely, how would you expect English security prices in Amsterdam to have changed relative to prices in London markets?
    b. When mail ship arrivals were delayed—for example, by four days—how would you expect English security prices in Amsterdam to have changed relative to prices in London markets?

CHAPTER

13

Trading Gone Awry

13.1 THE DARK SIDE OF FINANCE

When unfair or dysfunctional markets fail to draw traders, they cannot draw capital. The mere perception of dysfunction or unfairness is disruptive and can break down markets (Guiso, Sapienza, & Zingales, 2008). In Section 1.8, we discussed the importance and scale of financial trading to national and global economies. Trading does matter, and when trading goes awry, the ill effects can be far-reaching and devastating. In this chapter, we review a number of types of failures in trading along with their adverse effects on a variety of stakeholders, including traders, their employers, the financial services industry, and to society itself.

Securities fraud is expensive to securities firms, very difficult to detect, and more difficult still to evaluate, especially on an economy-wide basis. Consider the $139 billion in fines paid by the finance industry (these fines were not all connected to securities fraud) to U.S. enforcement agencies during the two-year period 2012–2014 and the many more fines that would have been assessed if not for failures of the affected firms or success in avoiding detection (Zingales, 2015). Dyck, Morse and Zingales (2014) note that the "main challenge in assessing the cost of fraud is that we only observe detected fraud: we do not know whether the observed fraud is the whole iceberg or just its visible tip."Based on a study of firms switching auditors after the demise of Arthur Anderson, the major accounting firm that failed in the wake of the Enron scandal, Dyck, et al. were able to estimate that only about a quarter of frauds are detected. Essentially, this study determined that Arthur Anderson detected only about a quarter of the financial frauds committed by their clients. The successor auditors to Arthur Anderson, who presumably had fewer disincentives to catch frauds, unveiled about three times as many frauds actually committed by their auditing clients. They extrapolated this "detection failure" rate to the economy. While most of the frauds relevant to this study were not trading frauds, the study might be suggestive of detection rates for various types of financial fraud.

Financial Trading and Investing
DOI: https://doi.org/10.1016/B978-0-323-90955-6.00013-6

© 2023 Elsevier Inc. All rights reserved.

The costs of securities fraud affect everyone. First, the difficult to observe indirect costs of securities fraud almost certainly exceed the direct costs. For example, securities fraud distorts investor capital allocation decisions. Gurun, Stoffman, and Yonker (2018) argue that investors significantly change their investment behavior as a result of a failure of trust in the financial services industry. In particular, they found that residents of communities with zip codes that were more exposed to the Madoff Ponzi fraud "withdrew assets from investment advisers and increased deposits at banks." The magnitude of this study's results (over $363 billion moved, 20 times the direct effect of the fraud itself) are large enough to imply that, overall, securities fraud has a significant impact on how investors allocate their capital. More generally, based on several extrapolations, Dyck et al. (2014) estimated the net annual costs of financial fraud at $380 billion, roughly a fifth of the market value of directly affected firms, and consistent with the proportional losses estimated by Karpoff, Lee, and Martin (2008). Dyck et al. further estimated that roughly one in seven firms engaged in fraud. Again, these frauds are not necessarily specific to trading, but trading fraud rates, extremely difficult to estimate, might be similar.

13.2 ILLEGAL INSIDER TRADING

Not all insider trading is illegal, either in earlier times or now. It will always be the case that some investors have better access to information than others. For example, sophisticated 19th-century English investors knew that the stock market tycoon Nathan Rothschild had pigeons and used them effectively to relay information. Then, as now, timely and accurate information was crucial to outperforming the market. In the early part of the 19th century, messenger pigeons were the fastest way to transmit news. For example, after the Battle of Waterloo, but before its outcome was publicly known in Britain, Rothschild's quiet and judicious (low-volume) selling activity was discovered by his countrymen who panicked and sold everything. However, Rothschild was merely manipulating the market. With prices down, and with a British victory over Napoleon at Waterloo, Rothschild actually bought huge quantities of stock at prices that his manipulative selling activity depressed. He then sold these shares at substantial profits when the news of the British victory finally reached England by more conventional news transmission.[1] Efficient acquisition and effective use of information along with appropriate manipulation of security prices were the trick to making stock market profits in the regulation-free 19th century. This case illustrates the importance of having better information than competing traders and acquiring the information sooner.

Illegal insider trading is traditionally defined as the execution of transactions on the basis of material nonpublic information in breach of a relationship of trust and confidence. We discussed this issue in Section 5.4. Further recall that there is no comprehensive statutory definition as to precisely what constitutes impermissible insider trading. Several statutes are generally relied upon to enforce prohibitions on insider trading including

[1] At least parts of this account might be apocryphal. Actually, while Rothschild did use courier pigeons to relay information, in this instance, his courier traveled by boat.

Section 10(b) of the Securities Exchange Act of 1934 and SEC Rule 10b-5. These rules are extended by *United States v. O'Hagan* to identify elements of a Section 10(b) or Rule 10b-5 insider trading claim, with elements of illegal insider trading summarized as follows:[2]

a. Possession of material nonpublic information
b. Trading while in possession of that non-public information
c. Violation of a relationship of trust and confidence (Flannery, 1998: 1–2)

In addition, Section 14 of the Securities Exchange Act and Rule 14e-3 impose a "disclose or abstain from trading" obligation on any person who trades in securities that will be sought or are being sought in a tender offer, while that person is in possession of material nonpublic information that he knows or has reason to know has been acquired directly or indirectly from the offeror, the subject corporation, any of their affiliated persons (including officers, directors, partners, or employees or any person acting on behalf of either company). Thus, the Securities Exchange Act of 1934 (Section 14e) provided for the initial federal statutory insider trading restrictions, although a 1909 Supreme Court decision in *Strong v. Repide* (discussed earlier in Sections 5.1 and 5.4) interpreted inside trading by a corporate director under special circumstances and with special information without appropriately revealing that information to be a form of fraud. Recall from Section 5.4 that *Chiarella v. United States* defined an insider as one who maintains a "relationship of trust and confidence with shareholders."[3] However, this definition can be rather broadly interpreted. As discussed in Section 5.4, the Insider Trading Sanctions Act of 1984 authorized penalties for illegal insider trading and the Insider Trading and Fraud Act of 1988 was intended to help define exactly what constitutes an insider and to extend penalties for illegal insider trading activity.

Notorious Insider Trading Cases

A casual web search for "insider tips" will lead to many sites selling information for sports betting, where, like the stock market, the best information is the key to success. However, selling and acting on insider trading tips in securities markets are illegal. Nevertheless, investors require the best possible information for their investment decisions, and at least a few cross the line into illegal trading. The following represent just of a few of the high profile insider trading cases pursued by the SEC and other regulatory authorities (plus one that predates the SEC and insider regulations). There are many more such cases pursued by authorities, and we can only guess as to how many that have not been pursued or even suspected (though, recall from Section 13.1 above efforts made by Dyck et al., 2014 to estimate financial fraud rates). One interesting aspect common to all of the individuals named in the cases that follow is that they were very successful in their businesses, and in most cases, quite wealthy—at least prior to the resolution of their cases.

[2]In an application of the "misappropriation theory," the U.S. Supreme Court in *United States v. O'Hagan*, 521 U.S. 642 (1997), held that a person can be held liable for violating Section 10(b) and Rule 10b-5 when misappropriating confidential information (in effect, stealing valuable information) for the purpose of trading securities.

[3]U.S. Supreme Court, *Chiarella v. United States*, 445 U.S. 222 (1980).

William Duer

William Duer was appointed Assistant Secretary of the Treasury in 1789 under Alexander Hamilton. Hamilton's and Duer's wives were cousins. Duer used information and connections obtained in his official position along with substantial leverage to speculate in new U.S. debt and various bank stock issues. He resigned his position in 1791 after learning that Treasury officials were to be prohibited from speculating in Treasury securities, but continued to use his inside connections for speculative purposes. His speculation and subsequent failure was a major cause of the Panic of 1792.[4] Geisst (2004) wrote that "[t]he New York City economy crashed along with him, and Duer was nearly disemboweled by an enraged mob that chased him through the streets. He died in debtors' prison a few years later."

Albert H. Wiggin

Albert H. Wiggin became chairman of Chase Bank in 1917, and engaged in M&A activities over the next 15 years, growing the bank to become the second largest in the U.S. Through 6 of his own corporations, including three in Canada to avoid reporting taxes on capital gains, Wiggin regularly traded Chase stock. During the Fall of 1929, he began short selling shares of stock of the company he ran, producing millions in profits for himself, a tidy sum of money at the time. The short selling of shares of his employer's stock broke no laws at the time, but the public certainly reacted badly to the clear conflict of interest, and his actions served as a rationale for the ban on insider trading provision of the Securities and Exchange Act of 1934.

Dennis Levine, Ivan Boesky, and Michael Milken

In 1985, Dennis Levine, a mergers and acquisitions specialist, was earning significant profits trading illegally on inside information through secret off shore accounts. He was a highly regarded managing director at Drexel, Burnham Lambert, once the fifth largest U.S. investment bank. Unfortunately for Levine, one of his trading account brokers, a Bahamian subsidiary of a Swiss bank, began "piggy backing" off his trades as did the bank's account executive at Merrill Lynch. Merrill Lynch became suspicious of these trades and notified the SEC. The SEC trail led back to Levine, who pled guilty in 1986 to securities fraud, tax evasion, and perjury. As part of his plea arrangement, he implicated fellow specialist Martin Siegel and famed arbitrageur Ivan Boesky, to whom he had passed illegal trading tips.

Boesky, considered the leading expert on merger arbitrage, had been paying Levine a percentage of his trading profits gained from Levine's tips.[5] Boesky agreed to a plea, whereby he would cooperate with authorities and secretly tape record conversations with his alleged confederates. Boesky, under the terms of his plea arrangement, implicated

[4]The formation of the NYSE shortly followed, partly in reaction to the Panic of 1792 and the related market manipulation, an example of an early effort at market self-regulation as the charter members of the new exchange realized that the market needed to be perceived as having integrity in order to survive.

[5]In a line later adapted by Gordon Gecko in the movie *Wall Street*, Boesky told a 1986 UC Berkeley Business School audience that "[Greed] is healthy. You can be greedy and still feel good about yourself."

Michael Milken and Drexel Burnham, then a major U.S. investment bank, along with Boesky's longtime friend, John A. Mulheren Jr., the former head of Jamie Securities. This led to the SEC launching an investigation into all of them.

Following up on his master's thesis at Wharton, Milken had innovated the 1980s junk bond market at Drexel. This led to extraordinary success for the firm, for which Milken was paid $550 million in 1987 alone. However, the investigation into the relationship between Boesky and Milken uncovered a $5.3 million payment in 1986 to Drexel that Boesky had characterized as a consulting fee. A significant amount of other evidence of insider trading was uncovered, but the bulk of the SEC's weak case against Milken and Drexel would be based on the testimony of Levine, Boesky, and other convicted felons. The SEC began to focus on Milken's brother Lloyd, who might have played a minor role in the insider trading activities (or at least would be subpoenaed to testify against his brother) and other relatives, including their 92-year-old grandfather. Milken agreed to a plea arrangement after the Department of Justice agreed not to pursue his brother. A codefendant of Milken, Alan Rosenthal, took his case to trial, whereby it was tossed by the judge before going to jury, lending significant credibility to the argument that the case against Milken was rather weak.

U.S. District Attorney Rudolph Giuliani threatened Drexel with an indictment under the Racketeer Influenced and Corrupt Organizations Act (RICO), which would have required Drexel post bond of $1 billion, a hefty amount for the heavily debt-financed firm. Even though company executives believed that the case against Drexel was weak, they decided that it would not be able to survive the costs and loss of reputation of a trial and agreed to pay a $650-million fine.

Levine received a two-year sentence, was fined $362,000, was disgorged of $11.5 million in illegal trading profits, and paid an additional $2 million in back taxes. Boesky served two years in Southern California's Lompoc Federal Prison, paid $100 million in fines, and was barred from the securities business for life. Milken received a 10-year sentence for fraud and illegal market manipulation, which was later reduced to 24 months; fined $600 million; and made restitution payments. Shortly after his release, he was diagnosed with terminal prostate cancer (but seemed much healthier 20 years later). Milken and other Drexel employees paid $1.3 billion into a pool to settle hundreds of lawsuits. Drexel was ultimately bankrupted in the aftermath of the scandal, fines, loss of reputation, and the collapse of the junk bond market.

Martha Stewart

Martha Stewart is a well-known television talk show personality and majority shareholder of NYSE-listed Martha Stewart Living Omnimedia, Inc. and of other companies. In 2001, she sold approximately 3928 shares of ImClone, which had just learned that its new prescription drug would not obtain Food and Drug Administration approval. After learning about the drug's setback, Sam Waskal, CEO of ImClone, quickly sold shares of his ImClone stock. His broker, Peter Bacanovic, who also served as Martha Stewart's broker, apparently notified her that Waskal was selling shares of his company's stock. The SEC investigated Bacanovic and Waskal, who were ultimately imprisoned. The SEC also began an investigation as to whether Stewart's selling was related to that of Waskal. Whether Stewart had engaged in illegal insider trading was not known at this point, and still is not

clear. Even if Stewart had been completely forthright about the information that she had received and why she sold, the SEC's case against her would have been weak at best. Unfortunately for Stewart, she conspired with Bacanovic to fabricate a story about a stop order at $60. That is, Stewart falsely claimed to have placed a standing order to sell ImClone shares once its shares dropped to $60. This fabrication led to charges of conspiracy, obstruction of an agency proceeding, and making false statements to federal investigators. She was convicted and sentenced in 2004 to serve five months in federal prison followed by five months of incarceration in her Bedford home (she was allowed to depart for up to 48 hours with an electronic monitor) and a two-year period of supervised release. She also paid a fine and disgorged the reduction in short-term trading losses that she avoided.

After she was charged in the alleged insider trading incident involving ImClone, but before the general public knew of the investigation, Stewart sold shares of her own company, Martha Stewart Living Omnimedia, Inc. This exposed her to additional insider trading charges, based on the negative impact that her legal problems might be expected to have on her company. She was charged for manipulating the price of her company's stock. These charges did not result in convictions.

Stewart's case is perplexing. First, she was a billionaire, and the stock sale saved her only $45,673 in losses. So, if she did sell illegally, her forgone losses were only a tiny fraction of her wealth. Worse, she almost certainly was no better off lying to the SEC than she might have been had she honestly admitted to the circumstances of her trade rather than concoct the story about the stop order. Better yet, she could easily have invoked her Fifth Amendment right to silence. The SEC's insider trading case against her probably would not have merited further action or at least not merited a conviction. In fact, she was not even indicted on insider trading charges in the ImClone case. Lying to authorities about the preexisting order to sell at $60 essentially exposed her to all of the charges on which she was convicted. Finally, Stewart did possess a reasonable degree of financial sophistication, having been a billionaire, a CEO, and a stock broker early in her career. She should have known better than to respond to investigators as she did. Perhaps she fell victim to a level of hubris that seems to affect some particularly successful people. Nonetheless, it is likely that neither her trading crimes nor the strength of evidence thereof seemed to warrant the sanctions that she suffered. On the other hand, the SEC and its crackdown efforts against illegal insider trading did receive substantial attention, a clear win for the SEC.

Raj Rajaratnam

In December 2011, Raj Rajaratnam, billionaire and founder of the defunct hedge fund, the Galleon Group, began serving an 11-year sentence in the federal prison system after being convicted on 14 counts of conspiracy and fraud. This was the longest sentence ever imposed for illegal insider trading, although less than half of the term requested by prosecutors. Rajaratnam was also fined $92.8 million and faced civil and other actions. Rajaratnam had been accused of having taken more than $50 million in insider trading gains based on tips provided by fellow conspirators at Goldman Sachs, McKinsey & Co., Intel, and Google, among others. Among those implicated in the investigations were Roomy Khan (Intel), Anil Kumar (McKinsey), Danielle Chiesi (Bear Stearns), Zvi Goffer (Galleon), Rajat Gupta (McKinsey and Goldman Sachs), Robert Moffatt (IBM), and the

extensive list continues. Many of these individuals have been charged, convicted, and served time in prison. Two interesting facets of this case were the large number of co-conspirators (over 60) and the use by authorities of wire and tapes to record conversations, including more than 40 wiretapped phone conversations between Rajaratnam and his accomplices.[6] Although pursuing illegal insider trading activities remains very difficult for authorities, its level of sophistication in the pursuit has improved. Furthermore, it seems interesting that such conspirator networks can be so large, given the conspirators' need to maintain secrecy.

SAC Capital

Perhaps due to the overlap of names that came up in the SEC's investigation of Galleon and Raj Rajaratnam, the SEC began a pursuit of SAC Capital Advisors, LP, a $15 billion hedge fund, and its founder, Steven A. Cohen, who had a personal wealth of approximately $9 billion. SAC was noted for imposing among the highest fees in the hedge fund industry, up to 50% of annual profits and 3% management fee versus the industry norm of 20% and 2%. Nevertheless, SAC averaged annual returns over 18 years of approximately 25% (United States v. S.A.C. Capital Advisors, L.P., 2013).

In August 2008, technology sector research analyst Jon Horvath, who pled guilty in 2012 to charges related to insider trading, sent his boss at SAC Michael Steinberg an email indicating that he had learned from someone at Dell that its upcoming earnings report would be below market expectations. Steinberg shorted shares in his portfolio and the email was forwarded to Cohen who, within 10 minutes of receiving it, closed his $12.5 million share position in Dell stock, thereby avoiding losses of approximately $1.7 million. Later in the month, Cohen complimented Steinberg, Horvath, and others in their group "Nice job on Dell." Cohen's defense on the insider trading charge was that he never read the email that he was forwarded prior to his sale of Dell.

According to the Department of Justice, Matthew Martoma, a portfolio manager at SAC had ongoing communications with Dr. Sidney Gilman, a neurology professor at the University of Michigan concerning an Alzeimer's drug, Bapineuzumab, under development at Elan and Wyeth. Gilman, who received $108,000 as consulting fees from SAC, and in violation of his duty to hold in strict confidence all information related to his participation in the clinical drug trial, provided Martoma with a summary of the drug's failing trial results prior to their public announcement. The day after a 20-minute telephone call between Martoma and Cohen, SAC sold approximately $700 million (including $400 million from Cohen's personal account) of Wyeth shares and Elan ADR shares and short sold an additional $260 million in shares just prior to the public announcement of the trial results, producing $276 million in profits and avoided losses—the largest insider trade series in history (as far as we know from the enforcement authorities). Gilman testified in exchange for avoiding prosecution. Martoma was rewarded with a $9 million bonus, but was ultimately convicted of conspiracy and securities fraud and sentenced to 9 years imprisonment.

[6]Transcripts of many of these conversations can be found on the web.

More generally, the Department of Justice argued that SAC was a hotbed of action for insider trading. It argued that traders were compensated on the basis of the quality of information provided to Cohen for his personal trading and were hired on the basis of their "edge," a presumed SAC euphemism for being able to produce inside information.

The SEC filed a civil action for insider trading against SAC in November 2012. In March 2013, SAC agreed to pay $616 million to settle charges of trading Dell, Elan and Wyeth. In July 2013, the S.E.C. instituted administrative proceedings against Cohen for failing to reasonably supervise Martoma and Steinberg. In addition, the U.S. Department of Justice filed criminal charges against SAC for wire fraud (one count) and securities fraud (4 counts), and against 8 former employees for insider trading. SAC pled guilty to all counts, agreed to stop managing funds on behalf of outside investors and pay an additional $1.2 billion, $900 million of which was a criminal penalty and the remainder as profit forfeiture. Five employees pled guilty to insider trading charges. Cohen, who was never personally criminally charged, later began operating another fund including approximately $9 billion of his own money and that of a small number of former SAC employees. As of 2021, he is running a new $20 billion venture capital fund, Point72 Asset Management, with roughly 1500 employees, as his ban on accepting outside money was lifted in 2015, and is the owner of the New York Mets baseball team.

Loeffler, Burr and Chris Collins

Despite passage of the 2012 STOCK Act (see Sections 12.4 and 12.6), allegations have been directed at more than 40 members of the U.S. Congress and the U.S. Senate for trading on inside information obtained in the course of their official duties and/or for failing to properly disclose their insider trades. For example, after a January 2020 confidential Senate briefing by the CDC director and Dr. Anthony Fauci in the early stages of the coronavirus pandemic, Senators Kelly Loeffler and Richard Burr made substantial trades on companies expected to be affected by COVID. Loeffler, wife of the CEO of ICE, the parent firm of the NYSE, and Burr, who voted against the 2012 ban on congressional insider trading, were strong supporters of the president and were quickly cleared by the U.S. Justice Department.

However, a more notorious case involved Representative Chris Collins, an early and vocal supporter of the then president, who was filmed in a White House lawn event on June 22, 2017, phoning inside information concerning Innate Immunotherapeutics to his son, Cameron Collins. Cameron sold 1,391,500 shares by the following Monday, June 26. Because the information was not obtained through his official duties, Collins was not charged under the STOCK Act, but for making false statements to the FBI and for trading on insider information obtained outside of his official duties. After repeatedly proclaiming his innocence, and narrowly being reelected to Congress, Collins resigned from his seat, pleaded guilty, and was sentenced to 26 months in prison. He received 7 months of COVID-related deferrals before starting his sentence in October 2020, was pardoned by the president, and was released two months later.

Monitoring Inside Trading Activity

It is usually very difficult for authorities to prosecute insider trading cases. Without substantial information from others—frequently codefendants—authorities often cannot

prove who knew what and when, or what has motivated profitable trades. Thus, successful insider trading prosecutions usually require cooperation from codefendants or leverage such as damaging information concerning other illegal behavior (such as tax evasion).

There have been recent improvements to enforcement efforts. First, surveillance techniques, which played significant roles in the Rajaratnam and SAC Capital cases, have improved considerably. For example, the SEC and all of the major markets and companies themselves are purchasing and developing software systems to monitor for illicit activity. The SEC is making use of varied resources to monitor insider activity, including formation of special surveillance teams, wire tapping, and bounty payments to informants. The SEC contends with roughly 700,000 tips per year from informants. Markets such as the NYSE, Nasdaq, and CBOE are making greater use of technology to monitor trading activity for suspicious activity. For example, FINRA, through its Office of Fraud Detection and Market Intelligence, has been using an intelligent surveillance application known as the Securities Observation, News Analysis, and Regulation (SONAR) system to detect suspicious trading patterns. Regardless, insider trading enforcement remains a challenging pursuit.

13.3 FRONT RUNNING AND LATE TRADING

Access to inside information clearly gives the trader a jump on the competition. So does access to information on other traders' activity, especially in the case of a broker with knowledge of her client's trades. Parasitic trading is the exploitation of knowledge of other traders' activities to enhance one's own trading profits.

Front Running and Parasitic Trading

Front running occurs when a broker uses her knowledge of a large pending order placed by his own client to buy (sell) a given security, and places his own order in front of the pending client buy (sell) order so as to benefit from the market reaction to the client's large order. Thus, a broker might illegally front run a large buy order by purchasing shares of stock for her own account after receiving a large purchase order for that stock from a client. Front running is an example of securities fraud because the broker has taken advantage of her client, in whose interest she is obliged to act. For example, by front running a client buy order, the broker's own buy order may force the stock price up against her client's buy order. In addition, the front running broker may also be in violation of insider trading law, in that she has nonpublic information about an order and violates her professional code of ethics by front running it. Front running can also occur when a broker passes on trade information to another client or trader.

In one well-publicized front-running scandal, the U.S. Attorney's Office and the SEC accused brokers from Merrill Lynch of allowing certain clients to listen to conversations involving other clients through broadcasts on Merrill Lynch internal speaker systems. In this "squawk box scandal," some clients were privy in advance to other clients' trades. Favored clients then rewarded the brokers with business, commissions, and cash. Several

former brokers from Merrill Lynch and Lehman Brothers were convicted and received probation or prison terms in 2010, as did several of their clients.

A similar, often legal, but frequently unethical practice is *tailgating*, a form of parasitic trading where the broker places an order immediately after the client's order. This practice is questionable because the broker is acting on the basis of the client's order, which itself might have been motivated by inside information or might precede additional client trading. Another type of ethically questionable parasitic order is *penny-jumping*, where the broker places a buy (sell) order one uptick above (downtick below) below the client's buy (sell) limit order, expecting to benefit either from the market's reaction to the client order or to limit his losses by transacting with the client. For example, if the client were to place a large buy order at 50.00, the broker would penny jump by placing her buy order at 50.01. The broker's order will execute ahead of the client's order due to price priority. If the client's order then executes, the broker will profit from any favorable price reaction due to the client's large buy order. If the price declines, the broker can quickly sell his shares to the client at a $0.01 loss.

Market Timing and Late Trading

We discussed in Chapter 4 the practice of *market timing*, which is the fund trading strategy intended to exploit stale security prices and NAV deviations from fundamental values. While market timing is not illegal, it can transfer wealth between fund shareholders and necessitate that funds maintain higher levels of cash to accommodate share redemptions. Market timing may increase fund transactions costs in response to shareholder redemptions. In addition, market timing could result in civil litigation due to prohibitions in the fund's prospectus or criminal penalties if a broker encouraged it to generate higher commissions for herself.

To limit these costs and problems, many mutual funds have announced in their prospectuses restrictions on the number of trades that customers may execute in and out of a fund per year. However, several scandals have erupted in recent years due to fund advisers having made undisclosed arrangements with favored customers to enable them to circumvent established trading restrictions. For example, in September 2003, New York State Attorney General Elliot Spitzer filed a complaint against the New Jersey–based hedge fund Canary Capital Partners, run by Edward Julius Stern, son of Leonard N. Stern. Spitzer's complaint alleged that Canary Capital entered into illegal agreements with mutual fund companies to defraud investors. For example, Canary made a secret agreement with Security Trust to enable it to engage in market timing-related trades. In another case, PIMCO agreed to permit Canary a specified number of trades in exchange for fees to be earned on $25 million it would manage for Canary.

In addition, Spitzer's allegations went further than these market timing agreement allegations, accusing Canary of *late trading*. In these allegations of illegal late trading, Canary maintained arrangements with funds to purchase or redeem shares after daily NAV's were computed. Unlike market timing, late trading is always illegal. Late trading was described by Spitzer as "betting on a horse race after the horses have crossed the finish line." According to Spitzer's complaint, Canary engaged in schemes to purchase mutual

fund shares after 4 PM exchange closing times with the assistance of certain Bank of America fund employees. Technically, Canary allegedly submitted trades in question prior to the 4 PM deadline so that they could be time-stamped before 4 PM. However, Spitzer alleged that the Bank of America employees would hold transactions until after the market closed, providing Canary the opportunity to "confirm" its trades. In fact, in 2001, Bank of America employees allegedly installed an electronic trading system in Canary's offices that allowed Canary to late trade by entering its orders directly, after the 4 PM deadline. Bank of America employees purportedly reconciled the "late" trades the following day. Canary Capital settled Spitzer's complaint for $40 million and was permitted to not admit guilt. Bank of America provided compensation to its mutual fund shareholders for related losses. Spitzer and the SEC also charged Janus, Bank One's One Group, and Strong Capital in other alleged incidences of late trading. In fact, the founder and chairman of Strong Mutual Funds, Richard Strong, was charged with late trading in his own fund.

13.4 BLUFFING AND MARKET MANIPULATION

The attraction of the securities markets to scoundrels can easily be likened to Jesse James's attraction to banks—*'cause that's where the money is*. Scam after scam has been created; most of the older ones seem quite naive to us. However, we have the benefit of hindsight. There has always been a sense among investors that market prices do reflect information—hence, investors rely on the market for quality information. If an unscrupulous operator is able to manipulate the market, he manipulates the information to which investors react. Much as the market creates information for traders, traders can, through their trading behavior, create information for other traders. If an investor has sufficient resources to create a given type of appearance in the market, he can generate huge profits through the trading activities of his counterparts. As we discussed in Chapter 5, most forms of manipulation in the information marketplace for securities are illegal.

Bluffing

In Chapter 6, we discussed how uninformed traders react to the information gleaned from the activities of informed traders. *Bluffing* is the act of fooling other traders into making unwise trades by convincing them that the bluffer has superior material information about security values. Bluffing can be a mechanism for manipulating the market. There are a number of ways that traders bluff, both legal and illegal. A trader can bluff the market by placing a bid or offer at a price or in a quantity that exaggerates his or her own true position, interest, or lack of interest in a security. For example, a bluffer might place a particularly attractive offer for a small quantity of shares, intending then to sell a much larger quantity of shares after a positive price response to the initial offer. As long as such orders are executed, they are usually perfectly legal. Bluffers may also simply spread rumors or false information about the value of a company's shares. Such false rumor spreading can represent deceit or fraud, illegal when used to manipulate markets, and perhaps otherwise as well.

Spoofing

Spoofing is the act of placing a quote that is intended to be canceled prior to its execution. Spoofing might be considered a form of bluffing when the trader places an order that she has no intent to execute. There are four potential motivations behind spoofing:

1. To overload the quotation system of a market, inhibiting its ability to execute trades. This disruptive activity is often called *quote stuffing*, and is usually illegal.
2. To delay a specific trader's trade execution by placing that trader's quotes behind a series of quotes with higher priority.
3. To create the appearance of false market depth or direction, particularly by submitting then later cancelling multiple bids or offers. This type of spoofing is often referred to as *layering*. Layering is frequently accomplished through a high-frequency trading (HFT) program designed to submit many quotations that are canceled after an executed transaction.
4. To create the illusion of liquidity, such as by rapidly repeating the transmission and cancellation of the same order, a practice known as *strobing*.

Clearly, spoofing can be used to manipulate a market in an illegal manner, and as we discussed in Chapter 5, has been specifically prohibited by Dodd-Frank. Orders, modifications, and cancellations are not considered to be illegal spoofing if they were submitted as part of a legitimate, good-faith attempt to execute a trade. Nevertheless, the rates of order cancellation reach over 95% in some markets, though, by no means are all of these cancellations related to spoofing. Order cancellations are a perfectly legitimate trading tool—when the original orders were placed in good faith, not when their cancellation was the original intent. Because it is difficult for authorities to distinguish good-faith order cancellations from spoofing, and because of the need to establish intent to defraud, authorities have faced significant difficulties with prosecution.

Damage to markets caused by spoofing can be far-reaching, as we will illustrate shortly in this section with respect to the May 6, 2010 "Flash Crash." Market participants lose faith in markets when they realize that prices and liquidity are illusory. Eggington, Van Ness, and Van Ness (2016) found that quote stuffing was pervasive, affecting approximately three-quarters of all U.S. listed stocks in 2010, and that intense episodes of quote activity impaired market quality, diminishing rather than improving liquidity and price discovery while increasing market instability.

Illustration: Spoofing

In 2001, the SEC alleged that Alexander Pomper submitted a phantom limit order (an order that Pomper did not intend to execute) on NASDAQ to purchase 300 shares of the thinly traded Gumtech International ("GUMM") at $11.375. This limit order improved the NBB by $0.3125 from $11.0625 to $11.375, while the NBO was $11.4375 per share. Pomper improved the bid for GUMM, but with a bid that he apparently did not actually intend to fill. Nonetheless, this bid did seem to improve the liquidity for GUMM, tightening the spread and making the stock appear to be worth more. Then, Pomper placed an order to sell 2000 shares of GUMM at $11.375 per share through another market-making firm. Pomper's sell order was immediately executed at $11.375 per share by this market maker

because the best (phantom) bid at $11.375 placed by Pomper seemed to suggest a higher value and more liquid market for GUMM. However, after executing its sell order, Pomper canceled his order to buy GUMM for $11.375, realizing additional profit of $625 ($2000 \times [\$11.375 - 11.0625]$) relative to the initial bid for $11.0625. In 2002, the SEC announced a settlement with Pomper, whereby without admitting or denying the allegations in the complaint (which also alleged other violations), Pomper agreed to pay $9800 in disgorgement and prejudgment interest and to pay a $15,000 civil penalty (Securities and Exchange Commission, 2002).

Spoofing in the Era of Dodd-Frank

As we discussed in Chapter 5, Regulation of Trading and Securities Markets, the Dodd-Frank Act specifically banned spoofing. Nevertheless, Michael Coscia, through his trading firm Panther Energy Trading LLC, placed a small bona-fide sell order and many large phantom buy orders at increasingly higher prices, creating the illusion of strong buying interest. When Coscia's small sell order was executed, he canceled his phantom buy orders, and then proceeded to implement the spoof in reverse order so as to execute a small buy order after creating the illusion of strong selling interest. These and related spoofing activities enabled Coscia to secure profits of approximately $1.4 million in 2011. These profits were later disgorged along with an additional $1.4 million fine in 2013, and Coscia was banned from trading for a year. In November 2015, the first criminal trial in the United States for spoofing led to guilty verdicts, for which Coscia was sentenced to 3 years imprisonment. In addition, the Chicago Mercantile Exchange and the U.K. Financial Conduct Authority imposed penalties on both Coscia and Panther.

Spoofing had long been considered a manipulative, deceptive and fraudulent practice under both the Securities Exchange Act and the Commodities Exchange Act, along with related SEC, CFTC and FINRA rules. However, since implementation of Dodd-Frank along with its specific reference to spoofing in 2010, the SEC, CFTC, FINRA and various exchanges have more actively pursued traders who engage in this type of activity. Among the many more recent targets of these regulatory bodies have been Coscia, Igor B. Oystacher and his trading firm 3Red Trading LLC ($2.5 million in penalties), Biremis Corp. (fined $250,000 and permanently banned from trading for charges related to layering), UBS ($700 million fine), Trillium Brokerage Services, and Navinder Singh Sarao (to be discussed shortly). In some respects, the HFT algorithms used in many spoofing schemes have been helpful in prosecutions in that their programming codes are used as evidence of intent to defraud; otherwise, proving intent can be quite difficult.

The Flash Crash

One of the more worrisome and perplexing short-term market crashes occurred on May 6, 2010. This "flash crash" had earlier been blamed on HFT. While HFT may well have contributed to the crash, its actual cause remains unclear. In addition to HFT, the May 6, 2010 flash crash has been blamed on fat fingers (keyboard errors, in this case, large accidental sell orders), technical glitches, debt default possibilities in Greece, etc. Regardless, the Dow Jones Industrial Average suffered its largest intraday loss ever, dropping by

almost 1000 points in less than 30 min before recovering almost as quickly. Prices of over 300 stocks and ETFs dropped by more than 60% relative to their prices only minutes earlier. Trades were executed at absurd prices, ranging from less than one cent to over $100,000 (We will discuss stub quotes shortly.). For example, some shares experienced odd gains, including Sotheby's, which increased in price from $34 to $99,999.99. Exchanges ultimately canceled over 20,000 executions.

As with many market crashes, it is difficult to pin-point the cause of the May 6, 2010 flash crash. There was a general concern that morning in the marketplace about whether the Greek government would default on its debt obligations. An early SEC report noted that the crash was preceded by a rapid algo-initiated short, reportedly by mutual fund group Waddell & Reed (not an HFT), of a $4.1 billion block of E-Mini Standard & Poor's 500 futures contracts on the Chicago Mercantile Exchange.[7]

However, HFT firms apparently were counterparties to these shorts, and immediately covered themselves by further shorting (turning the contracts into hot potatoes, discussed shortly), drying up liquidity in these markets. Within 20 minutes of the start of the "crash," trading in the E-mini contract was halted for 5 seconds, which allowed the market to recover. When contract trading resumed, prices quickly recovered most of their losses.

In 2015, 5 years after the flash crash, the U.S. Department of Justice filed fraud and market manipulation charges against Navinder Singh Sarao, an independent trader who operated from his parents' home near London. Mr. Singh had engaged in spoofing activities, having used his computer-based algorithms to place thousands of orders for E-Mini Standard & Poor's 500 stock index futures contracts. His algorithms had canceled over 19,000 orders by the afternoon of May 6. While it is doubtful that this trader or any one of the issues described above was the sole cause of the flash crash, each might have played some contributing role.

Buy, Lie, and Sell High

"Pump-and-dump" schemes occur when the market manipulator touts company's stock with false and misleading statements to the marketplace, causing the stock's price to rise before selling his own shares at inflated prices. Although pump and dump schemes predate the telephone by more than 200 years (e.g., the South Sea Bubble), the traditional misinformation campaign is launched through a boiler room (telemarketing) operation such as those conducted by Stratton Oakmont in the 1990s. Stocks promoted by the schemes tended to be thinly traded shares of small companies. More recent campaigns have exploited improved mass communications technology, including e-mail, social media, investment research websites and newsletters, hacked websites, and Internet chat rooms.

For example, 15 years ago, it had been estimated that as much as 65% of e-mail traffic was unsolicited (spam), and that 15% of this spam was touting shares of stock (Frieder & Zittrain, 2008). Such spam operations have been linked to pump and dump operations, where operators tout shares of stock and then sell them. Frieder and Zittrain claim that stock volume for touted shares increases dramatically after spam is delivered. Frieder and Zittrain describe one scenario:

[7]See the joint report of the Commodities Futures Trading Commission and Securities and Exchange Commission (2010).

In a well-known case brought by the U.S. Securities and Exchange Commission, stock touter Jonathan Lebed routinely purchased stock accounting for anywhere from 17% to 46% of the stock's market volume for a day, and sent spammed e-mail touts on the same day. He then lodged limit orders to sell for the next day's trading session, anticipating a rise in the stock price after the general public received his touts and some acted on them. Lebed's case and subsequent settlement with the S.E.C. focused on his failure to disclose his own financial interest in the securities he touted; as discussed below, stock touts today often include such disclosures (See also Securities and Exchange Commission, 2000).

Frieder and Zittrain found that on days prior to touting, and on days touting takes place, returns are positive. Returns after touting are negative. Two-day returns average −5.25%, worsening further when the intensity of touting increases. Surely no reader of these results would consider purchasing stock on the basis of unsolicited spam. Yet, someone must be. Who? Why? And why does this buying behavior continue to persist even after transaction returns are so consistently poor?

Jonathan Lebed, the stock touter described above, was a 15-year-old trader from Cedar Grove, New Jersey. In September, 2000, the SEC settled 11 cases of stock market fraud against this high school student that had resulted in gains ranging from $12,000 to $74,000 from September 1999 to February 2000. With interest, the SEC disgorged $285,000 in illegal profits. With his AOL connection and hundreds of Yahoo Finance postings under numerous fictitious names, Jonathan had apparently purchased shares in small companies and then posted numerous buy recommendations under fictitious names, increasing share volume from 60,000 to over a million per day in the affected companies. He also maintained a website, stock-dogs.com, where he published his opinions and recommendations. It was reported that Jonathan's father, Greg Lebed, had a heart attack and other heart issues during the period of Jonathan's trading and SEC battles. However, the ending of this story wasn't all bad for Lebed. In his settlement with the SEC, Lebed was permitted to keep over $800,000 in profits from other transactions. For years, Lebed has remained an active trader, distributed a newsletter, and maintained a website touting penny stocks.

Banging the Close

The CFTC defines *banging the close* as a "manipulative or disruptive trading practice whereby a trader takes positions in a large number of futures contracts during the closing period of a futures contract (that is, the period during which the futures settlement price is determined) in order to benefit an even larger position in an option, swap, or other derivative that is cash settled based on the futures settlement price on that day."[8] Consider the case involving the giant (now defunct) hedge fund Amaranth Advisors, LLC, which incurred $6.4 billion in losses in the early spring 2006. The fund's failure arose from losses in trading highly leveraged natural gas contracts on NYMEX. The CFTC later charged Amaranth and its former head trader, Brian Hunter, with trying to manipulate natural gas futures prices. More specifically, Hunter was accused of banging the close and flooding the then-open outcry NYMEX futures markets with "series of rapid and successive" orders for natural gas during the final minutes of trading before the close, thereby forcing prices down in this less active market as buying interest diminished. The actual

[8]See http://www.cftc.gov/ConsumerProtection/EducationCenter/CFTCGlossary/glossary_b.

intent of his selling activity, according to the CFTC, was to depress natural gas prices in the ICE, where Amaranth held much larger short positions through swap contracts.

Corners and Pools

Corners involve the purchase a sufficient level of a given security to obtain market power over its price. The purpose of the corner is to manipulate the price of the security, an activity made illegal by the 1934 Securities and Exchange Act. Numerous corners have existed over the years in commodity and securities markets. For example, the Hunt brothers (Nelson and Lamar), two of the then-wealthiest men in the world, attempted to corner the silver markets in the early 1980s, taking billions of inherited dollars in long positions. Although they succeeded in bidding up the price of silver almost 10-fold through their purchases, the bottom fell out of the market before they could unload their holdings. They nearly went bankrupt. A *pool* is a common fund comprising participants who combine their resources to obtain a large position in a security in order to manipulate its price. Pools might be used to facilitate cornering of a market, and were a common activity prior to the 1929 stock market crash.

Vanderbilt v. Drew

One of the most interesting cases of market manipulation involved Commodore Cornelius Vanderbilt, a 19th century tycoon, and Daniel Drew, an operator with connections to Tammany Hall. As of the early 1860s, Vanderbilt had owned enough shares (with a cost as low as $8 apiece) of Harlem Railroad Company stock to effectively control the company and its share price, bidding it up to $100. He intended to extend the railroad's service into Manhattan and required approval from the city to accomplish this. However, Daniel Drew, also a Harlem Railroad investor, sold his shares and shorted 137,000 shares, almost five times the number of outstanding shares. He used his influence to convince Boss Tweed to repeal the railroad's license to operate in New York City, with the intended effect of ruining the railroad. The price of shares fell from $100 to $72, with Vanderbilt purchasing shares as Drew sold. In fact, Vanderbilt had, in effect, purchased the company several times over by acting as counterparty on Drew's short sales, and had the market cornered. In time, the short selling peaked and Drew needed to cover his short positions by repurchasing the stock. But only one major investor owned shares—Vanderbilt. Vanderbilt, having cornered the market, set his price at $179 per share, a price calculated to be the maximum that Drew could afford to pay. Drew was forced to settle with Vanderbilt at his terms, Vanderbilt obtained his approval to extend the railroad into Manhattan with Drew's cooperation, and Drew was practically bankrupted in the process. Drew had fallen victim to a market *corner*, intended to *squeeze the short seller*[9]:

> He who sells
> What isn't his'n
> Buys it back
> Or goes to pris'n.

[9]This quote has been attributed to Drew.

Modern Day Corners

Such corners are not confined to "ancient" history. For example, two brokers employed by the now defunct firm Moore & Schley, Cameron & Company were accused in the late 1980s of acquiring shares of Chase Medical Group for the purpose of artificially inflating their prices.[10] In effect, these brokers were accused of cornering the market for these shares and engaging in phony transactions to prop up the share prices.

In the height of the financial crisis of 2008, the German automobile manufacturer Porsche revealed that it owned 42.6% of the stock of Volkswagen, along with call options on another 31.5% of its shares. A number of hedge funds had short sold Volkswagen stock, believing that it would drop once Porsche reached its target ownership level of 75%. Unfortunately, the state of Lower Saxony in Germany owned another 20% of the shares, which it did not want to sell, creating a shortage of shares in the market available for short sellers to cover their positions. This shortage of available shares caused the share price to rise from about 200 Euro to over 1000. In effect, without breaking German securities law, Porsche had, in effect, cornered the market for Volkswagen shares and squeezed hedge fund short sellers, earning significant paper profits on its shares.

Market corners are not confined to equity markets. The U.S. Treasury maintains a strict set of rules for trading its instruments intended to sustain a perception of fairness. For example, no bidder is permitted to bid for or obtain 35% or more of a given Treasury issue. These regulations are intended to prevent market corners. Nonetheless, it still might be possible to abuse trading and bidding rules to create an unfair advantage. For example, in summer 1991, the venerable investment bank Salomon Brothers created a huge scandal by routinely bidding on more than 35% of issues, illegally bidding on as much as 105% of certain new issues. To conceal their actions, many of their bids were placed under names of clients who were unaware of the activity. Bonds were "parked" in client accounts. Such overbidding enabled Salomon to corner the market and fix prices for new two-year treasuries, enabling the firm to generate substantial trading profits by engaging in short squeezing.

The Treasury responded to this scandal by expanding its list of primary market participants and changing bidding procedures. Salomon fired its CEO, John Geutfrend, and several key employees including John Meriwether and Paul Mozer, the primary instigator of the fraud.[11] Mozer subsequently pled guilty on two counts of lying to the Federal Reserve Bank of New York in his bid submissions, served four months in a minimum security prison, and was fined $30,000 and an additional $1.1 million in an agreement with the SEC that barred him from the securities business for life. Warren Buffet, who owned a substantial stake in the firm through his holdings in Berkshire-Hathaway, contributed significant credibility to the firm by taking over as interim CEO and is often credited with saving the firm, which since consolidated with Smith-Barney into CitiGroup.

[10]See Aggarwal and Guojun (2006) and SEC, Litigation Release No. 11973, January 25, 1989 for details on this case.

[11]We discussed Meriwether before in Chapter 7 with respect to his leadership of the failed hedge fund Long-Term Capital Management.

Wash Sales

Unscrupulous market participants have always had an interest in manipulating market conditions to generate profits. For example, consider *wash sales*, which are sham transactions intended to create the appearance of sales where, in effect, no sales actually take place. The SEC defines a wash sale as a transaction that involves no change in beneficial ownership. Wash sales may be intended to manipulate security prices (e.g., conspirators execute transactions with one another to create records of sales prices to deceive other participants in the market). One may arrange to sell a security to a cohort at a fixed (as opposed to market-determined) price. The transaction and its price are revealed to the market, which is intended to accept this price as the market-determined value of the security. A series of wash sales at ever-increasing prices, known as a *jitney game*, can give other investors the impression that the value of the security is rising, leading investors to purchase the stock. The market manipulators benefit when investors are led to purchase from them the security at inflated prices. It is quite likely that the security's price will drop rapidly once investors realize that they had been tricked into purchasing the security. The market manipulators hope that they will no longer own the security when this happens. Wash sales may also involve the sale of securities for tax purposes with offsetting transactions to repurchase the securities or related instruments. The IRS defines a wash sale to be a sale and a re-purchase of the same security within 30 days. This sort of wash sale is not illegal if not intended to manipulate the market, though any capital losses and associated tax write-offs generated by this activity will be disallowed by the IRS.

Fishing

Fishing refers to a practice intended to obtain secret or hidden information concerning order sizes in dark pools. Fishing occurs when a trader (who might also be called a "gamer" in this setting) sends a series of small orders to a dark pool to detect whether there is a large order waiting in that pool. If the fishing results in successful order fills, the trader might infer that additional orders or a very large order await in the pool. Then, the trader can submit a much larger order at a size and price that reflects the information that he has obtained through his gaming activities. Suppose, for example, that an institutional trader has discretion reflected in his algo to buy shares at a price as high as 50.03 and places a bid at 50.00. Suppose that the current market for the shares is 50.00 bid/50.05 offer. Now, suppose that another trader, a prospective "fisher" or "gamer", detects this algo order, but does not know its reserve price to purchase shares. The gamer "pings" the algo by offering shares at 50.04. If the offer is not immediately executed, it immediately cancels. Then, the gamer offers 50.03 and the institutional algo buys. The gamer now realizes that it has found the institution's reserve price for the shares. The gamer now knows that it can probably sell many more shares at 50.03 rather than the 50.00 quoted bid. Thus, the gamer submits a much larger sell order at 50.03. This sort of gaming activity is often perfectly legal. In some cases, the gamer can manipulate the selling price of shares by entering small buy orders in the displayed market to force the displayed price up, say from 50.05 to 50.07. This might force the institutional seller to raise its bid to 50.05, at which time the gamer sells larger numbers of shares at this higher price.

As we discussed in Chapters 4 and 5, an important exception to the quote rule requiring that all market centers publicly disseminate their best bids and offers through the securities information processors allows an exception for quotes that are immediately executed or canceled, such as fill or kill orders. These quotes are often not publicly displayed, but may be displayed to select clients. These orders (see discussion from Section 4.9) can result in flash trading in advance of the public quote stream. Such flash orders can facilitate fishing and have been criticized as being very unfair to the majority of market participants.

Predatory algo trading strategies are designed to exploit other institutional algo orders. Suppose, for example, the predatory trader detects an institutional algo order to buy a large number of shares. Suppose that a small number of these shares are offered at 50.00, and the predatory trader suspects that the institution has pegged its order to the NBBO and is willing to pay as much as 50.25 per share. The predatory algo can seek to lock in a profit by artificially increasing the share price. Upon detection of this institutional order, the predatory trader places a small bid of 50.01, then continues placing small bids at successively higher figures as the institutional algo trader increases its bid with the NBBO. Ultimately, the predatory algo places a short sell order at 50.25, knowing that it created much of the apparent demand running the price up to 50.25. When the price falls, the predatory algo covers its short position.

The Odd Eighths Scandal

In Section 5.7, we introduced a somewhat different type of market manipulation, perpetrated by conspiracy involving Nasdaq dealers, sometimes referred to as the "Odd-Eighths Scandal." Quotes data collected during the 1990s revealed that Nasdaq dealers quoted prices far more often in even-eighths than in odd-eighths, artificially widening spreads by using round dollar amounts, 1/4, 1/2, and 3/4 fractional amounts (Christie & Schultz, 1994). This thoroughly widespread manipulation in Nasdaq markets enabled dealers to pay less for and sell securities at higher prices, enhancing profits, and enhancing commissions in agency transactions. The ensuing S.E.C. investigation led to much negative publicity for Nasdaq, including revelations of "Chinese markets" and harassment of or refusal to trade with brokers who were uncooperative by offering deals that were "too good" to their clients or counterparties.[12] At the time, Nasdaq was not required to post public limit orders, so that the best prices would not have been known by or available to retail customers. This S.E.C. investigation eventually led to censure of Nasdaq and fines exceeding $1 billion. In 1997, the S.E.C. imposed new rules requiring that brokers and dealers post limit orders and their best prices on Nasdaq as well as in ATS markets such as Instinet.

[12]The term "Chinese market" used by conspirators referred to a "disorderly market" in which dealers competed on the basis of price; Nasdaq markets at the time were largely insulated from competition. An August 11, 1996 *Wall Street Journal* article (p. 3-3) compared the Chinese market to a "Chinese fire drill" where teenagers would leap out of their car at a stoplight and jump back in at a different door.

LIBOR and Fraudulent Trading

The London Interbank Offer Rate (LIBOR, introduced in Chapter 2) derives from a daily survey of 18 global banks originally conducted and compiled by the British Bankers' Association (BBA), a UK lobby group, and reported by Thomson Reuters with respect to rates at which the banks believe they can borrow a "reasonable" number of dollars (and certain other currencies) from one another in the London interbank market. Actually, there are a number of LIBOR figures, for different maturities and currencies. LIBOR has been important because it is widely used as a benchmark for interest rates on which many loans and securities are based. A trader able to influence this rate or even rates reported by just one of the banks participating in the survey should be able to obtain significant profits trading products based on the LIBOR. This is a conspiracy scheme that involved roughly 30 traders and other employees at Rabobank Coöperatieve Centrale Raiffeisen-Boerenleenbank B.A. (Rabobank). Similar schemes were undertaken at Barclays, UBS, Citigroup, JPMorgan Chase, and RBS. These schemes ultimately led to major upheavals in the process of obtaining and computing rates for the LIBOR, which is now managed by the Intercontinental Exchange, the parent firm of the NYSE.

Essentially, one such scheme involved Rabobank employees responsible for submitting Rabobank rates for the BBA surveys communicating with their own traders concerning their positions in LIBOR-related products, and these traders requesting that the survey submissions to the BBA be biased so as to favor their own LIBOR-related positions and anticipated trades. Web chat and email trails reveal that traders requested these accommodations, rate-submitters complied with the requests and that the LIBORs were biased as a result. Anthony Conti, a senior trader on the money markets desk for Rabobank, and Anthony Allen, the bank's global head of liquidity and finance in London, were, respectively, sentenced to 24 and 12 months in 2015. They and a number of other employees were fined, and Rabobank paid over $1 billion in fines to the U.S. and U.K. However, the criminal convictions against Conti and Allen were overturned on appeal in July 2017 on Fifth Amendment grounds, because evidence against them was obtained from testimony the men were forced to give in the United Kingdom, which does not offer the same Fifth Amendment protections as does the U.S. A witness against them in their U.S. trial had also been compelled to provide testimony in a related case in the U.K.

This LIBOR scandal had several serious ramifications. The over $9 billion in fines assessed to banks and the criminal prosecutions were really just a start. The takeover by the ICE Benchmark Administration from the BBA was a direct result of the scandal. Perhaps the most important effects of the scandal concern the erosion of trust in the LIBOR by regulators and markets, and the erosion of public trust in the market more generally. Use of LIBOR has been in the process of being phased out, with a current target date set for June 30, 2023, for many applications, and generally being replaced with the Secured Overnight Financing Rate (SOFR; see Chapter 2) where practical. LIBOR had served as the benchmark for over $300 trillion in loans worldwide and many times this amount in various derivative contracts. Many existing contracts have needed to be rewritten to account for the LIBOR phase-out.

Other Quote Abuses

Quote matching occurs when a small trader places a quote one tick from that of a large trader so as to profit from the large trader's transaction price pressure, or to use the large trader as a counterparty should prices reverse. Quote matching exploits the option associated with a limit order. See end-of-chapter exercise 12 in Chapter 6 for an illustration of quote matching.

Quote stuffing, discussed earlier in this section above, is the placement of large numbers of rapid-fire stock orders, with most or all of which being canceled almost immediately, frequently for the purpose of clogging HFT and other algorithms and data computations. The purpose of quote stuffing is not to execute orders, but to sabotage algorithms of other traders. See the description of spoofing above. More generally, most types of spoofing are intended to deceive other traders, quote stuffing is intended to clog the trading system and/or sabotage other traders' algorithms. Be aware that the two terms are sometimes used synonymously. Both types of activities are usually illegal.

Many dealers are obliged to post quotes for even their least liquid securities. They sometimes fulfill their obligations by posting quotes that are very distant from any reasonable market price for the security (e.g., bid for $0.01 or offer at $100,000), not intending or expecting that these quotes would ever be executed. However, in a one-sided market, these *stub quotes* might result in executions as liquidity dries up, causing securities to trade at absurd prices. In many instances, the trades and profits realized with stub quotes were entirely unintended, but executed automatically by algorithms or due to stale quotes. Nevertheless, stub quotes have the potential to destabilize markets and increase volatility.

13.5 PAYMENT FOR ORDER FLOW

Payment for order flow, a topic that we introduced in Sections 1.6 and 4.8, was originated by Bernard Madoff in the 1990s. During that era, stock prices were quoted in eighths of a dollar, considerably more than the $0.01−0.02 per share that he offered retail brokers such as Charles Schwab and Fidelity to route retail order flow to him. Madoff reasoned that the retail clientele of such firms tended to be uninformed noise traders, hence in no position to exploit inside or special information as were insiders or the informed traders that we discussed in Chapter 6. He was happy to trade against and to pay for order flow that originated only with uninformed traders; he faced no adverse selection problem with them. The S.E.C. allowed this practice to continue, reasoning that its potential for conflicts of interest would be mitigated by competition as other firms began to engage in the practice. Furthermore, it was not clear at that point that retail clients were being hurt as they still tended to trade at prices close to or at the NBBO.

Nonetheless, there has also been significant controversy over these rebating practices as it has since spread to other market makers, smaller exchanges, and then to the larger exchanges. When exchanges and their competitors sell virtually identical products, they compete for customer order flow, usually on the basis of prices or trading costs. Presumably, this competition enables customers to obtain better prices. However, many brokerage houses internalize orders and own or make arrangements with alternative

trading system (ATS) markets. This means that brokerage firms have a special incentive to internalize orders or send their orders to these ATS markets where possible—the firm will receive commissions on the trade, earn trading profits, receive payment for order flow, and secure additional business for their own systems. Clearly, many investors trade through brokers with strong potential conflicts of interest. In fact, markets even pay brokers to direct their orders to them so as to enhance their liquidity and the profitability of their traders and market makers. In addition, markets usually have order imbalances, which means that certain additional orders will serve to create liquidity while others will consume liquidity in that market. It might make sense for such markets to charge higher prices on orders that consume liquidity than those that create liquidity.

This system, also known as *make-or-take pricing*, refers to exchanges allowing patient traders to post standing limit orders that await execution until some other trader takes the other side. In effect, the patient trader is a market maker who receives payment from the exchange for making the market (liquidity rebate) and the market taker pays an access fee. Note the connection between this system and the "immediacy argument" of Demsetz described earlier in Section 6.5. These rebate programs have the potential to help customers get price improvement for their orders by creating competition among the exchanges, market makers, and liquidity providers with whom they trade, potentially resulting in superior execution quality. In addition, the practice has virtually eliminated commissions and fees for many investors. A number of observers have noted that make-or-take pricing systems have reduced spreads and improved liquidity, although the liquidity rebates normally are not reflected in bid and offer quotations. This practice continues to grow more widespread, and since February 2010, even the NYSE, which had been the most vocal opponent of payment for order flow, has been making such payments to providers of liquidity.

On the other hand, such payments and rebating for order flow might lead to market abuses, reduce price competition for securities, and reduce security price transparency. At a minimum, the practice creates a conflict of interest for brokers; it is not clear that the rebate system provides for the best executions, at least for retail traders. For example, FINRA fined Robinhood $1.25 million in December 2019 for failing to ensure that its customers received the best prices for their orders. Battalio, Corwin, and Jennings (2016) suggest that payment for order flow leads brokers to send orders to venues that might be inferior with respect to fill rates and other indicators of execution quality. In addition, routing orders to dark trading venues inhibits price discovery in markets, obscuring what might be the best executions for retail traders. Furthermore, the maker-taker fee structure makes it difficult to determine the best price as required by Reg NMS because displayed quotes do not reflect rebates or fees. The customer might lose the opportunity to obtain an unannounced better price from a broker on the national exchange floor. Regardless, in 2000, the SEC adopted Exchange Act Rule 11Ac1-6 (now SEC Rule 606), requiring broker-dealers to make available quarterly reports that present a general overview of their routing practices. These quarterly reports must identify the significant venues (exchanges or ATS markets) to which customer orders were routed and disclose the material aspects of the broker-dealer's relationship with such venues. In addition, the Rule requires broker-dealers to disclose, on customer request, the venues to which the customer's individual orders were routed.

13.6 FAT FINGERS, HOT POTATOES, AND TECHNICAL GLITCHES

In 2005, a novice trader in a subsidiary of Japan's second-largest bank, Mizuho Financial Group, attempted to sell one share of J-com stock on the Tokyo Stock Exchange for ¥610,000 ($5041). By accident, he transmitted an offer to sell 610,000 shares at ¥1 each. In effect, he attempted to sell $3.075 billion worth of stock for $5041. These 610,000 shares that he attempted to sell represented 41 times the total number of J-com shares issued by the Japanese recruitment company. Mizuho attempted to cancel and reverse the order, but not quickly enough, so that this incident (losses exceeding ¥27 billion) wiped out Mizuho's entire profit for the quarter. The absence of filters either at the bank or at the Tokyo Stock Exchange to prevent such an obviously erroneous trade played havoc with the market for J-com stock, and caused the entire Nikkei-225 to drop by 1.72%. The broader market reaction was largely in response to traders attempting to guess which firm had made the mistake, driving down share prices for nearly all Japanese financial institutions. Later litigation assessed a significant portion of the blame on the Tokyo Stock Exchange, and Mizuho recouped some of its losses.

Fat fingers (keyboarding and other entry errors) appear to have led to many isolated incidents of sudden security price movements. For example, three years before the Mizuho incident, a trader at Bear Stearns seemed to have caused a 100-point drop in the Dow after inadvertently entering a $4-billion sell order instead of his intended $4 million buy order. Morgan Stanley made a similar mistake with a 2004 order for $10.8 billion rather than the intended $10.8 million. Another incidence involved Diebold Inc., which increased by 30% in six seconds on June 2, 2010 on NASDAQ and BATS exchanges. Fat fingers can have significant consequences. Washington Post Co. dropped by 99% in less than one second on June 16, 2010 on NYSE Arca, and Progress Energy Inc. increased by 90% in less than one second on September 27, 2010 on NASDAQ. Nikkei 225 futures contracts dropped by 1.1% on June 1, 2010 as a result of an unintentional algo order placed on the Osaka Stock Exchange. Each of these occurrences illustrates the need (and failures) for effective trade filters, even in the absence of automatic algo executions. However, even an effective trade filter would not have helped the Bank of America trader who set up a $50 million trade on his computer while awaiting instructions from his supervisor on whether to execute. While he waited, one of his colleagues tossed a rugby ball that landed on his keyboard, causing the order to execute before it was authorized.

Technical Glitches

Despite many efficiencies arising from technologies for trading and markets, various computer, software, and routing malfunctions cause regular and significant disruptions in markets and trading firms. For example, on December 9, 1987, a squirrel burrowed through a power line in Trumbull, Connecticut where Nasdaq's data facilities were located. The squirrel's adventure halted Nasdaq trading for 82 minutes and got himself killed. A similar event occurred in 1994, halting trading on Nasdaq for 34 minutes, and killing yet another squirrel.

As the securities industry further automates, more frequent technical glitches occur. For example, consider the period 2012–2013. Nasdaq's computer systems were overwhelmed during the Facebook IPO on May 18, 2012, when a surge in order cancellations and updates delayed the opening of the shares for trading and impaired later trading. Nasdaq settled with the SEC, agreeing to pay $10 million for failing its "obligation to ensure that their systems, processes and contingency planning are robust and adequate to manage an IPO without disruption to the market" (SEC 2012) and settled a class action suit with Facebook investors in 2015 for $26.5 million.

Perhaps even more disruptive was the approximately $460 million in losses incurred by the Knight Capital Group, resulting from its accidental acquisition of approximately $3.5 billion in net long positions in 80 stocks and approximately $3.15 billion in net short positions in 74 stocks. Knight's problematic launch of insufficiently tested software led to over 4 million erroneous executions and many more trade cancellations over 45 minutes on August 1, 2012. The new software aggressively executed transactions at a rate of $2.6 million per second, causing significant slippage, raising prices as it bought and reducing them as it sold. NYSE markets were disrupted and affected stocks experienced enormous price volatility. Knight, which had been the largest equities trader in the U.S., was fined $12 million by the SEC, struggled to recover and was taken over by Getco to form KCG in 2013.

In August 2013, Goldman Sachs unintentionally flooded U.S. exchanges with erroneous stock-option orders, losing roughly $100 million. Two days later, a connectivity issue between a trading firm and the UTP SIP (Nasdaq's securities information processor) impaired the ability of the SIP to disseminate quotes and trades and caused a three-hour halt in trading at Nasdaq. Such technical glitches by America's best-known investment bank and major stock exchange do little to instill confidence in technology-based trading systems, and imply the possibility of disastrous market crashes. But they also should serve as warning to traders: Expect the unexpected. Better yet, prepare for the unexpected.

Hot Potato Volume

One of the most important functions of a securities market is to convey information through the price aggregation process. Trading conveys information about securities, and higher volumes are often thought to convey more or higher quality information. However, hot potato trading, repeated passing of inventory imbalances among dealers, seems to actually dilute the information content of trading (Lyons, 1997). This type of trading tends to be more common in the presence of HFTs and where interdealer volume is high. On October 4, 2010, Gary Gensler, Chair of the CFTC referring to the May 6, 2010 flash crash, remarked that:

> Much of the volume on May 6 [the Flash Crash] was just positions being moved back and forth over a matter of seconds between high-frequency traders and other market makers. This is what our economists refer to as "hot potato volume." For the large trader's order to actually be absorbed by the market, it had to find fundamental or opportunistic buyers who were willing to hold the position at least for more than a few seconds (Gensler, 2010).

An important lesson from this and other mini flash crashes is that high trading volume and market liquidity are not the same, especially when hot potato volume is involved.

A Textbook for $23,698,655.93 (Plus $3.99 Shipping)

A trading algo program gone bad might seem more understandable when you consider Peter Lawrence's 1992 developmental biology textbook, *The Making of a Fly: The Genetics of Animal Design*. Michael Eisen, an evolutionary biologist at the University of California, reported that he sent one of his post-docs to purchase a copy on Amazon.com on April 18, 2011, who was shocked to see that Amazon listed 17 copies for sale: 15 used from $35.54, and 2 new from $1,730,045.91 (+$3.99 shipping) (See Rooney, 2011). The two new copies were offered by two booksellers, Profnath and Bordeebook. Their prices quickly increased from over $1.7 million to $2.8 million, and by the end of the day, to $3,536,675.57. Ultimately, the book's price topped out at $23,698,655.93, plus shipping.

How did this price manage to evolve to over $23 million? First, Profnath used an algorithm to undercut Bordeebook's price, at 0.9983 times Bordeebook's. However, Bordeebook used its algorithm that would set its own price at 1.270589 times Pronath's, apparently seeking the larger markup. This was the recipe needed for the book price algorithm to explode: competing booksellers' price multiples exceeding one.

Such algo programs gone bad are not confined to markets for books, and they are not one-time isolated events. Consider the case of Hansen Transmissions (HSNTF.PK), a company that manufactures and supplies wind turbine gearboxes. On January 23, 2009, its stock price increased from $1.62, the prior close, to $143.32 on the trading of 100,000 shares, apparently on the news of an alternative energy speech given by President Obama. The next day, trading opened at $1.69. While this scenario itself was not a major market calamity, it does warn of potential meltdowns (e.g., the May 6, 2010 "flash crash") or other market disruptions that might occur with unfiltered or even filtered algo trading.

13.7 ROGUE TRADING AND ROGUE TRADERS

Markets can be perceived as unfair for many reasons. Some investors might have better access to information than others. Certain investors are able to exercise market power due to their size, access to information, domination, or transaction volume. As we have discussed in Chapters 4 to 6, investors lacking these advantages might be unwilling to trade with their advantaged counterparties; such perceived unfairness can lead to market failure. This chapter is largely concerned with the impact that unfair trading activities can have on traders, markets, and institutions. However, illegal trading activities are often less significant causes of market failure than trading misguided by poorly designed compensation and managerial control systems. This scenario seems most apparent with failures associated with the financial market crises of the early 21st century. Perhaps, more than any other concept, the following scenarios illustrate *Darley's law*, named for John Darley, a professor of psychology at Princeton University who emphasized the importance of implementation and design of effective risk management systems to accompany objective incentive–based compensation systems:

> The more any quantitative performance measure is used to determine a group or an individual's rewards and punishments, the more subject it will be to corruption pressures and the more apt it will

be to distort and corrupt the action patterns and thoughts of the group or individual it is intended to monitor.[13]

Rogue trading is usually characterized as systematic unauthorized trading, trading with unapproved counterparties, or trading of unapproved products. Rogue trading normally involves traders exceeding risk limits and/or loss limits set by their employers and is accompanied by efforts to conceal unauthorized actions from management and regulators. That is, rogue trading is often used as a cover for prior mistakes. In practically all instances, rogue trading is accompanied by poor internal control systems and supervisory failures that allow these deceptive actions to succeed. Rogue trading, in and of itself is not necessarily illegal (laws can be pretty specific here), it is by definition, not authorized by management. However, rogue trading is often accompanied by illegal actions to camouflage the unauthorized trading. Rogue traders do not normally face criminal charges due to trading; their illegal deceptions to cover their inappropriate trading lead to criminal prosecution. In many instances, rogue trading is motivated by what behavioral economists call an *irrational escalation of commitment*, which might be likened to "throwing good money after bad."

The SEC, FINRA, CFTC, and exchanges require broker-dealers to supervise their employees and maintain compliance departments as part of an integrated risk management system to protect the institution and its clients, yet almost all rogue trading can be tied to breakdowns or lapses in supervisory control systems. Trading firms need to implement supervisory control systems designed to detect and deter all unauthorized trading, including those that have strong profit potential. Warning systems for limit breaches should be among the first line of defenses against rogue trading. Employees should not have access to reporting systems used to monitor their activities, a rule regularly broken in the rogue trading scenarios discussed below. Red flags suggesting the need for further questioning and/or examination should include any discrepancies in reports of account activity, the presence of any unauthorized products in trading accounts, traders working unusually long hours and working into the late night, unusually large unrealized profits and losses on unsettled transactions, movements of positions between accounts especially prior to reviews, and unusual patterns in and large numbers of cancellations and corrections. Policies mandating vacations extending at least two consecutive weeks during which traders are denied all access to trading, reporting and control systems are useful for detecting unauthorized trading. Paying particular attention to the activities of traders who request exceptions to such policies including requests to relax controls is useful as well. Lapses in essential principles such as these are important warning signals of the potential for rogue trading.

Rogue Traders

Grant & Ward

News of one of the earliest documented rogue trading disasters broke in 1884, engineered primarily by Ferdinand Ward, a habitual con man. Ward had recruited Ulysses S. Grant, Jr.,

[13]Darley (1994), perhaps based in part on Campbell's law (Campbell, 1976) on "high-stakes" testing in the educational system.

to form the brokerage firm Grant & Ward, with his father, the former U.S. president Ulysses Grant, later serving as an unwitting silent partner. Ward used the former president to falsely claim that he had inside information concerning government contracts. The firm made bad trades, illegally rehypothecated securities already serving as pledge for other purchases on margin, altered the firm's books, and for a time, covered liquidity needs by using a Ponzi scheme (discussed in the next section [13.8]) that promised investors 10% per month. Liabilities of the firm ultimately exceeded assets by a factor of more than two to one. The firm failed, Ward went to prison, the former president was bankrupted, and the Wall Street Panic of 1884 ensued, worsening the economy already mired in a depression.

Nick Leeson and Barings Bank

One of the most notorious rogue traders in history was Nick Leeson, chief derivatives trader at Barings' Singapore office. (See Waring and Glendon, 1998, for details.) In 1995, Leeson's trading activity singlehandedly brought down the centuries-old Barings Bank, forcing the sale of Britain's oldest bank to ING for £1. The proud bank, which had long served the English "upper crust," financed Napoleonic Wars, the Erie Canal, and the Louisiana Purchase, had been founded in 1762. Until its takeover, it had served as Queen Elizabeth's personal bank. Leeson incurred £880 million in trading losses, sufficient to bring down the entire bank. Leeson was authorized by the bank to execute futures and options orders for clients or for other units of Barings Bank and to arbitrage price differences between Nikkei futures traded on the SIMEX and the Osaka Exchange. Both were inherently low-risk activities and did not provide much latitude for exposing the bank to higher-risk speculative activity. However, to camouflage poor trading performance and his losses, Leeson created a fictitious trading account numbered 88888 (8 is a lucky number in Chinese numerology) into which he recorded fictitious trades offsetting his own trades. Leeson claimed that this account was initially created to hide an error made by a colleague; rather than purchase 20 contracts as a customer had ordered, the colleague sold them, costing Barings £20,000. Regardless, Leeson used this account to hide his own bad trades that followed.

When Leeson was discovered, he escaped Singapore to Malaysia aboard his yacht. The world followed as he then escaped to Thailand and then Germany, adding international intrigue to his story. Leeson was caught and served three years in a Singapore prison before being released for good behavior. He has since become somewhat of a celebrity ex-rogue trader, giving lectures to companies on risk management. He sold his story to a newspaper and, while in prison, wrote the book *Rogue Trader*, which became the basis for the movie of the same title starring Ewan McGregor. After being awarded time off for good behavior in prison, he went on to serve as the CEO of the Galway United Football Club in Western Ireland, to maintain an active schedule on the lecture circuit, and to write his second book, *Back from the Brink: Coping with Stress*.

Clearly, Barings' control systems were lacking. More than 20 of Barings' employees were implicated in failing to detect the fraud. Barings management was warned about the failure to segregate back office and trading operations, but they failed to implement remedial actions. Essentially, Barings failed to implement effective compliance mechanisms, proactive risk management measures, appropriate oversight, and sound management oversight.

What are the lessons to be learned from Leeson's fraud? The bank neglected the first and most important step in implementing risk controls: sufficient segregation of trading

and back office (record-keeping) functions. This should be the first step in preventing and stopping trading fraud. In effect, Leeson was his own supervisor. Second, management did not react appropriately when it was warned of increased concentration of financial risks from a relatively small trading unit. A substantial proportion of this U.K. bank's profits were generated by a small trading unit in Singapore. In fact, higher-level managers might have intentionally overlooked this and other warning signs. Third, continued trading fraud, at least initially, is often committed not for personal gain, but to cover poor performance or losses. Each escalation is often an attempt to cover the previous failure. Fourth, incentive-based compensation contributes to trading fraud, not only motivating the perpetrator, but his supervisors as well. In fact, incentive-based compensation systems might have greater effects in failed supervisory systems than in rogue trading activity itself. Fifth, a single rogue trader can bring down even the largest and most venerable of financial institutions. Finally, the superstar trader should merit the closest—not the least observation, not only to ensure that his trading is legitimate, but to better understand the secrets to his success.

John Rusnak and AllFirst

John M. Rusnak was a foreign exchange (FX) trader for the Baltimore-based AllFirst Financial Corp., which was owned by Allied Irish Banks PLC (AIB). Rusnak gambled incorrectly in 1997 that the U.S. dollar would fall against the Japanese yen. He used currency forward contracts to purchase yen as the dollar strengthened, inflicting several millions of dollars in losses on his employer. He hid these bad trades by attempting to trade out of them with larger trades, a doubling-up tactic used by many rogue traders. Second, Rusnak manipulated internal controls and entered fictitious trades and data into AllFirst's books and records, hiding the facts that he exceeded his trading limits and assumed excessive risks. Again, his activities illustrate the failure of financial institutions to maintain effective and independent control systems. In an effort to cover up his losses, Rusnak created a fictitious businessman with a mail drop at Mail Boxes Etc. in New York to confirm bogus currency trades with AllFirst's independent auditors. Over the next four years, as his trading losses mounted, Rusnak, who projected himself as a church-going family man, hid his losses with his fictitious account and other manipulations of his employer's control system. He wasn't caught until December 2001, by which time his losses had exceeded $691 million. Like Leeson, he disappeared and was later caught after his losses were revealed. Rusnak was not charged with stealing money, although he did earn $650,000 in bonuses related to his trading activity, along with a modest salary. He pled guilty to bank fraud on October 24, 2002, and received a 7-1/2-year jail term followed by five years of supervised probation, along with a $1000 per month restitution obligation (57,000 years of years of payments are required at this rate). Earlier in 2002, AIB had sold majority control of AllFirst to M&T Corporation. Some of the same lessons from the Leeson case apply to Rusnak, except that from Rusnak, we learn that massive compensation levels are not a necessary condition for rogue trading.

Orlando Joseph Jett and Kidder Peabody

Orlando Joseph Jett, an MIT graduate with a Harvard MBA, began trading strips (zero-coupon Treasury instruments) and other Treasury instruments for the well-regarded firm

Kidder Peabody in 1991, failing to report significant profits in his "rookie" year and earning a bonus of $5000. Trading profits recorded by Jett in 1992 and 1993 were $32 million and $151 million, followed by $81 million in the first quarter of 1994. These apparent trading profits were realized on trading volumes of $25 billion, $273 billion, and $1.567 trillion for 1991, 1992, and 1993, and $1.762 trillion for the first 3 months of 1994. In 1993, Jett was named Kidder Peabody's "man of the year" and was awarded a $9 million bonus. His compensation awards were not merely a fixed proportion of his profits; compensation proportions were increased as his trading profits increased. Furthermore, Jett's supervisor, Ed Cerullo, was awarded a $20 million bonus in 1993 for Jett's accomplishments. However, it was later revealed that Jett did not actually generate profits for his employer in those years; his desk actually lost $85 million. The "phantom" profits resulted from a mistaken entry in Kidder's internal accounting system where bond payments to be made later were not discounted; that is, future cash flows were being treated as though they were immediately realized. Ultimately, when the contracts settled, Kidder Peabody realized the losses, but Jett was able to cover these losses (on paper, with the help of the flawed internal accounting system) by increasing the size of his positions, realizing more phantom profits. Interestingly, Jett was ultimately cleared of criminal wrongdoing because it could not be proven that he knew about and intentionally exploited the glitch in the accounting system, even though he repeated transactions with almost identical results and with increasing frequency. In fact, Jett argued that Kidder ordered him to increase his risks to deceive Kidder's parent, GE, into believing that the Kidder unit was realizing its profit objectives. When GE became concerned about the ballooning sizes of Jett's trading volume, Jett was ordered to cut back on his holdings. He did so, but apparently entered offsetting transactions to exploit the faulty accounting system and to avoid having his bonuses cut. Only when his positions were cut back were his trading losses uncovered. Jett was ultimately fined $200,000 in 2007 and ordered by a judge to repay $8.2 million in losses. Kidder Peabody was sold to Paine Webber in 1994 shortly after the scandal broke, which later merged into UBS.

A significant contributor to Jett's losses at Kidder Peabody seems to have been lack of proper oversight by his supervisor, Ed Cerullo, who earned substantial bonuses based on Jett's phantom trading profits. Cerullo had a strong incentive to not adequately supervise his employee. Might Cerullo and his superiors have made a conscious decision to tolerate Jett's ballooning trading volume and probable risk limits simply because of his previous "successes" and their ballooning compensation levels? Might supervisors and higher-level managers have been complicit in the rogue trader's seeming success? A second contributor to many trading unit failures is that trading losses can often be covered with taking positions with increased risks, a cycle that can play out until the entire firm fails. Third, as we saw in the Leeson case, close monitoring of unusually high trading profits is probably as important as close monitoring of any unusual trading activity.

Jérôme Kerviel and Société Générale

In 2008, Jérôme Kerviel of Société Générale confessed to a €4.9 billion fraud where he misappropriated computer access codes, falsified documents, and employed other methods to cover his improper trading activities. It was the largest discovered trading fraud in history until Bernard Madoff's detection (which was actually more of reporting and misallocation frauds than a trading fraud). Kerviel's case was particularly interesting because of

how ordinary and unimpressive he seemed in almost all respects. Unlike Leeson and Jett, Kerviel was by no means a trading superstar. He apparently made very profitable, but unauthorized trades that he hid by intentionally executing transactions to create losses. He argued that the type of behavior that he engaged in was commonplace in trading units, but that employers turned a "blind eye" to these activities as long as the activities appeared to be profitable. Kerviel was formally charged in France in 2008 with abuse of confidence and illegal access to computers. He was ultimately sentenced to five years in prison (two years suspended) and to make full restitution of the lost €4.9 billion, though the restitution was overturned on appeal. He was also given a permanent ban from employment in the financial services industry although in 2016, he won a €450,000 judgment from Société Générale on the grounds that his employer was aware of his deceitful behavior, and, therefore, his dismissal was unfair.

Potentially Authorized Trading Gone Wild

However, not all "trading gone wild" constitutes unauthorized rogue trading, though authorized large-scale trading and poor risk controls can be just as damaging. Consider former highly compensated JPMorgan Chase trader Bruno Iksil, also known as the "London Whale," and his team of traders who lost approximately $6 billion trading credit default swaps and other instruments, partly due to inadequate risk controls. It appears that bank supervisors may have been aware of the large and risky trade positions. The bank was criticized by regulators for its high-risk trading strategies, weak management, poor response to this crisis, and failing to co-operate with regulators. Ultimately, the bank agreed to pay approximately $920 million in fines to U.S. and U.K. regulators to settle charges related to this "London Whale" trading debacle, accepting responsibility for engaging in unsafe and unsound practices and failing to oversee its traders. In its investigation of this fiasco, the U.S. Senate's Permanent Subcommittee on Investigations found that Chase "disregarded multiple internal indicators of increasing risk; manipulated models; dodged oversight; and misinformed investors, regulators and the public about the nature of its risky derivatives trading."

13.8 HIDING TRADING LOSSES WITH PONZI SCHEMES

Ponzi schemes are not trading schemes. However, Ponzi schemes have been used by rogue traders to mask illegal or unprofitable trading activity such as the illegal rehypothecation of securities in the Grant & Ward disaster discussed in the previous section. Each of the Ponzi schemes described here share an essential quality: They promised high returns with little risk. This is typical. More generally, red flags that an investment opportunity might actually be a Ponzi scheme include promises of high returns with little risk, remarkably stable investment returns over time, unregistered sellers or securities, and difficulties withdrawing money from the investment. In the case of Grant & Ward discussed in the previous section, the Ponzi scheme covered trading losses, magnified by excess margin and illegal securities rehypothecation, which continued until the firm imploded.

In December 2007, one of the highest profile and most respected members of the Wall Street community, Bernard Madoff, revealed to his son that his investment firm, Madoff

Securities, was a Ponzi scheme. Essentially, this meant that the Madoff lost or pocketed clients' money, and when asked or forced to meet client obligations, used funds raised from other clients to meet these obligations. This amounts to a sort of pyramid scheme to "rob Peter to pay Paul." Although Madoff was a trader and operated a well-respected trading firm, his Ponzi scheme was not a trading fraud per se. Madoff used the Ponzi scheme to cover his trading fraud. Essentially, Madoff lied to his investing clients about trades that he executed along with the profitability of those trades, and then used his Ponzi scheme to cover his trading fraud and associated losses, and to keep clients temporarily satisfied.

After the confession, Madoff's son promptly turned his father in to authorities, and investors in his fund realized that their losses totaled $50 billion to $65 billion (including profits that were never realized), far and away the largest Wall Street fraud incident ever. Madoff pled guilty to 11 counts of securities fraud on March 12, 2009. Previously, Madoff had maintained a highly successful trading and market-making business, accounting for as much of 12% of Nasdaq volume. He was an active securities market regulator, having served on the NASD and NASDAQ boards during much of the 1980s, and had even served in 1990–1991 as chairman of the board of Nasdaq. However, the fraud occurred in a business distinct from his trading operation, his investment advisory operation. Madoff accepted large sums of money from "feeder funds" such as Walter Noel's Fairfield Greenwich Group, Tremont Capital, Stanley Chais, and Avellino and Bienis as well as banks such as Santander. Madoff prohibited these funds from listing his fund in their prospectuses and was able to attract over 3000 investment clients without registering with the SEC as an investment advisor (registration is required when the advisor's client base exceeds 15). This enabled Madoff to conduct his advisory operations outside of the purview of the SEC. Actually, the SEC had reviewed Madoff's operations repeatedly, overlooked the Ponzi scheme evidence, and focused on whether Madoff's investment advisory operation was obliged to register with the SEC. Madoff actually used the SEC reviews to bolster his credibility, maintaining that the SEC had regularly reviewed his operations and found nothing amiss.

Madoff regularly provided statements and trade confirmations to clients, apparently almost entirely fabricated. He claimed to be engaging in a trading practice known as "split strike conversion," a simple collar strategy involving the S&P 100. Madoff claimed to take long positions in a basket of stocks resembling the S&P 100, purchasing S&P Index 100 puts while writing S&P calls. In a Black-Scholes framework, his strategy should have produced a return comparable to the riskless rate, much less than the 14% to 20% that he actually "paid" to most of his investors. David Friehling, whose tiny accounting firm Friehling & Horowitz in Rockland County, New York, audited Madoff's records and was paid approximately $12,000 to $15,000 per month (after 2004), was also charged with securities fraud.

Carlo Ponzi

The Ponzi scheme as we discuss it in this book is not so much a trading activity as it is an activity intended to cover up another crime. For example, Madoff used it to cover his failure to produce promised investment gains and to support his fraudulent reporting. The Ponzi scheme is named for an Italian immigrant to the United States, Carlo Ponzi,

who in 1919 discovered an arbitrage opportunity for which he required financing. He had spent most of his working life as a dishwasher, waiter, clerk, and so on, but learned that he could trade postal coupons obtained for one cent in Spain for six cents in postal coupons in the United States. Hence, he proceeded to solicit investors to participate in this opportunity, promising returns of approximately 50% in 45 days. The scheme seemed simple enough and thousands of investors participated. Several government agencies examined Ponzi's trading operations, but found no illegal activity. Over several months, thousands of investors demanded their money back. Ponzi complied with each demand, enhancing his credibility. Essentially, each investor was paid from proceeds realized from the sale of securities from other investors and not enough investors demanded money back to render Ponzi insolvent. Payment from new investor proceeds was necessary because the red tape and various obstacles and expenses made the original arbitrage scheme unprofitable. When a Boston newspaper questioned the legitimacy of Ponzi's operations, state regulators investigated Ponzi's books and prohibited him from selling additional securities. Ponzi was still able to meet demands for cash, further enhancing his credibility. He invested company proceeds into a new bank, the Charles Ponzi Company, which brought in even more money. In time, bank auditors declared the bank to be bankrupt and it was revealed that Ponzi had served time in prison for an earlier fraud. This caused his operations to collapse and created widespread panic and even caused a number of other banks to fail. Ponzi was jailed for mail fraud and spent the remainder of his life in and out of prison for various types of fraud before dying in Brazil without enough money to be buried.

Other Ponzi Schemes

Bernard Madoff's fund is by no means the only investment company that has turned out to be a Ponzi scheme. For example, a 2009 complaint filed by the SEC alleged that from at least February 1995 to 2009, Joseph S. Forte of Philadelphia had been operating a Ponzi scheme (Securities and Exchange Commission, 2009a). Apparently, Forte fraudulently obtained approximately $50 million from roughly 80 investors through the sale of securities in the form of limited partnership interests in his firm and codefendant, Joseph Forte, L.P. Forte, who never registered with the SEC, apparently told investors that he would invest the limited partnership funds in a securities futures trading account in the name of Forte LP that would trade in futures contracts, including S&P 500 stock index futures. However, he continually lost money on the limited trading that he did, sustaining trading losses of approximately $3.3 million between 1998 and 2008. From the inception of the scheme, the SEC alleges that Forte and his firm lied to investors about the returns on the trading, reporting annual returns ranging from 18.52% to almost 38%. In addition, the SEC alleged that Forte stole significant portions of the funds entrusted to his care.

James Ossie of Atlanta was charged by the SEC along with his firm CRE after raising at least $25 million from over 120 investors offering "30-day currency trading contracts" that guaranteed a 10% return in 30 days (Securities and Exchange Commission, 2009b). In addition, the SEC alleged that CRE and Ossie claimed to have generated profits sufficient to pay these guaranteed returns by trading U.S. and Japanese currency

contracts as the exchange rate fluctuated. Apparently, Ossie told investors that his program involved very little risk because CRE had established a large, defensive reserve fund from which to pay back the 10% return on investment along with any principal that investors chose to redeem. The SEC alleged that Ossie and CRE never generated sufficient returns from currency trading to pay the promised returns, but was merely a Ponzi scheme.

Sir R. Allen Stanford and the Stanford International Bank based in St. Croix sold certificates of deposit from Antigua that offered "unusually high and consistent returns" (double the market average, according to *Business Week*). C.A.S. Hewitt, the tiny Antiguan accounting firm that audited Stanford's books, was run by Celia Hewett, who went missing at about the time the scandal broke and Charlesworth "Shelly" Hewett, who died in early 2009. Stanford "managed" $51 billion, slightly more than the amount that Madoff managed, and seemed to be financing a lavish lifestyle on a Ponzi scheme. Stanford was included on the 2008 Forbes list of billionaires, owned an island in the Caribbean and a number of jets, helicopters and yachts. In early 2012, Stanford was convicted on 13 counts of fraud, conspiracy, obstructing justice, and violating U.S. securities laws, and was sentenced to 110 years imprisonment.

Additional Reading

Chapter 2 of Kolb and Overdahl (2006) provide a nice discussion of market abuses and regulation in the futures markets. Lefevre's 1923 classic *Reminiscences of a Stock Operator* remains an entertaining and informative read, and is still the most widely referenced book on stock market scams. For a read on trading gone awry on an economy-wide level, MacKay's classic 1841 *Extraordinary Popular Delusions and the Madness of Crowds* is probably the most important reading material for bubbles and crashes during the first half of the 19th century. The 2001 *New York Times Magazine* article by Michael Lewis on the teenager Jonathan Lebed is a thoroughly entertaining read. Shiller's 2000 best-seller *Irrational Exuberance* remains a readable, relevant, and important book. Michael Lewis's *Liar's Poker: Rising through the Wreckage on Wall Street* (1989) and *The Big Short: Inside the Doomsday Machine* (2010) are entertaining reads on the Wall Street mindset. James Stewart's 1992 classic *Den of Thieves* remains among the most thorough and readable books on 1980s-era inside trading scandals. Finally, several relevant movies might be fun to watch. *Liar's Poker*, *When Genius Failed*, *Den of Thieves*, *Rogue Trader*, and *Wall Street* all are worth a watch. The staff report prepared by the Security and Exchange Commission (2021) summarizes some very recent market updates, including new brokerage firm marketing efforts, zero-commission transactions and fractional share offerings along with certain undesirable market activities, including short squeezes.

References

Aggarwal, R. K., & Guojun, W. (2006). Stock market manipulation. *Journal of Business*, *79*(4), 1915–1954.

Battalio, R., Corwin, S. A., & Jennings, R. (2016). Can brokers have it all? On the relation between make-take fees and limit order execution quality. *Journal of Finance*, *71*(5), 2193–2238.

Campbell, D. T. (1976). *Assessing the impact of planned social change*. Dartmouth College: The Public Affairs Center.

Christie, W. G., & Schultz, P. H. (1994). Why do NASDAQ market makers avoid odd eighth quotes? *Journal of Finance*, *49*(5), 1830–1840.

Commodities Futures Trading Commission, Securities and Exchange Commission. (2010). *Findings regarding the market events of May 6, 2010: Report of the staffs of the CFTC and SEC to the Joint Advisory Committee on Emerging Regulatory Issues*, September 30.

Darley, J.M. (1994). *Gaming, gundecking, body counts, and the loss of three British cruisers at the Battle of Jutland: The complex moral consequences of performance measurement systems in military settings*. Speech to Air Force Academy, April 6.

Dyck, A., Morse, A., & Zingales, A. (2014). *How pervasive is corporate fraud?* Working paper No. 2222608. Toronto: Rotman School of Management, University of Toronto.
Eggington, J. F., Van Ness, B. F., & Van Ness, R. A. (2016). Quote stuffing. *Financial Management*, *45*(3), 583–608.
Flannery, A. (1998). *Insider trading cases: Settlement criteria and recent developments*. Paper presented at Institute of Continuing Legal Education in Georgia, Fourth Annual Securities Litigation and Regulatory Practice Seminar, October 8.
Frieder, L., & Zittrain, J. L., (2008). *Spam works: Evidence from stock touts and corresponding market activity*. Berkman Center Research Publication No. 2006-11(2007), 13 Hastings Communications and Entertainment Law Journal 479.
Geisst, C. R. (2004). *Wall Street: A history from its beginnings to the fall of Enron*. New York: Oxford University Press.
Gensler, G. (2010). *Remarks, swap execution facility conference*. Washington, DC, October 4.
Guiso, L., Sapienza, P., & Zingales, L. (2008). Trusting the stock market. *Journal of Finance*, *63*(6), 2557–2600.
Gurun, U. G., Stoffman, N., & Yonker, S. E. (2018). Trust busting: The effect of fraud on investor behavior. *The Review of Financial Studies*, *31*(4), 1341–1376.
Karpoff, J. M., Lee, D. S., & Martin, G. S. (2008). The cost to firms of cooking the books. *Journal of Financial and Quantitative Analysis*, *43*, 581–612.
Kolb, R. W., & Overdahl, J. A. (2006). *Understanding futures markets* (6th ed.). New York: Wiley-Blackwell Publishing.
Lefevre, E. (1923). *Reminiscences of a stock operator*. New York: George H. Doran Company.
Lewis, M. (1989). *Liar's poker: Rising through the wreckage on Wall Street*. New York: W.W. Norton.
Lewis, M. (2001). Jonathan lebed: Stock manipulator, S.E.C. nemesis—and 15. *New York Times Magazine*. February 21. Available at <http://www.cs.brown.edu/people/rbb/risd/Lebed.html> Accessed 25.08.09.
Lewis, M. (2010). *The big short: Inside the doomsday machine*. New York: W.W. Norton & Co.
Lyons, R. K. (1997). A simultaneous trade model of the foreign exchange hot potato. *Journal of International Economics*, *42*, 275–298.
MacKay, C. (1841). *Extraordinary popular delusions and the madness of crowds*. New York: Crown Publishers.
Rooney, B. (2011). For sale: The making of a fly, only $23,698,655.93 (Plus $3.99 Shipping. Tech Europe: The Wall Street Journal April 26). Available at <https://blogs.wsj.com/tech-europe/2011/04/26/for-sale-the-making-of-a-fly-only-23698655-93-plus-3-99-shipping/> Accessed 15.10.17.
Securities and Exchange Commission. (2000). Administrative Proceeding File No. 3-10291, September 20.
Securities and Exchange Commission. (2001a). Administrative Proceeding File No. 3-10631, November 5.
Securities and Exchange Commission. (2001b). SEC charges six individuals with spoofing, SEC Release 2001-129, Washington, DC, November 5.
Securities and Exchange Commission. (2002). Securities and exchange commission v. Alexander Mark Pomper, Civil Case Number CV-01-7391, SEC Litigation Release No. 17479, Washington, DC, April 19.
Securities and Exchange Commission. (2009a). Securities and exchange commission v. Joseph S. Forte, et al., Civil Action No. 09-0063, SEC Litigation Release No. 20847, Washington, DC, January 8.
Securities and Exchange Commission (2009b). Securities and exchange commission v. CRE Capital Corporation and James G. Ossie, Civil Action No. 1: 09-CV-0114, SEC Litigation Release No. 20853, Washington, DC, January 15.
Securities and Exchange Commission. (2012). *SEC charges NASDAQ for failures during Facebook IPO*. <https://www.sec.gov/news/press-release/2013-2013-95htm> Accessed 06.30.2017.
Securities and Exchange Commission. (2021). Staff report on equity and options market structure conditions in early 2021. Available at: <https://www.sec.gov/files/staff-report-equity-options-market-struction-conditions-early-2021.pdf> Accessed 12.06.2021.
Shiller, R. J. (2000). *Irrational exuberance*. Princeton, NJ: Princeton University Press.
Stewart, J. (1992). *Den of Thieves*. New York: Simon & Schuster.
Strong v. Repide. (1909). Justia US Supreme Court Center– 213 U.S. 419 (1909). Supreme.justia.com. <http://supreme.justia.com/us/213/419/> Accessed 20.08.12.
United States v. S.A.C. Capital Advisors, L.P. (2013). Southern District of New York no. 13-CV-5182.
Waring, A., & Glendon, A. I. (1998). *Managing risk*. London: Thomson Learning.
Zingales, L. (2015). Presidential address: Does finance benefit society? *Journal of Finance*, *70*(4), 1327–1363.

13.9 EXERCISES

1. Describe why it is difficult to manipulate security prices in a well-developed and liquid market.
2. **a.** Most markets, including those for labor, commodities, and real estate, allow insider trading activity. Transactions in these markets are routinely based on unequally distributed information. Why should insider trading be common and legal in these markets?
   b. Discussion in Section 13.2 focused on the damage to markets done by insider trading. Prepare an argument to justify legalize the practice of insider trading.
3. Does insider trading violate the Securities Act of 1933? If so, what section? If not, why not?
4. In a release dated November 5, 2001, the SEC announced an action against Israel Shenker. In this announcement, the SEC provided the following description of a series of transactions executed by Shenker:

   > 14:44: 44—Shenker placed an order to buy 500 shares of the target security at $34.0625, and directed that it be routed to the ECN. The order raised the NBBO bid price from $31.50 to $34.0625.
   >
   > 14:44:46—Shenker entered a limit order to sell 500 shares of the target security through SOES at a price of $34.0625, and the order was immediately executed.
   >
   > 14:44:48—Shenker canceled his buy order, causing the NBBO bid price to drop from $34.0625 to $31.25.

   By manipulating the public quote to obtain a better execution price for his 500-share sell order, Shenker was unjustly enriched in the amount of $1281.25 (Securities and Exchange Commission, 2001a, 2001b).

   a. What is the name of the type of market abuse alleged by the SEC?
   b. Why were these manipulations not considered to be wash sales?
   c. Why was the activity alleged by this announcement illegal?
5. In the chapter, we discussed a number of problems in markets that arise from legal practices and norms, but might be worthy of discussion by regulators. Consider the following issues:
   a. Payment for order flow and the liquidity maker-taker pricing schemes have been criticized for creating conflicts of interest between brokers and their clients. Devise and discuss some strategies for mitigating the conflicts that are associated with these practices, assuming that the practices are continued.
   b. Traders have complained that market liquidity is often not what it appears because when they place market orders in response to quotes; the quotes disappear before their orders are executed. That is, quotes are canceled between the times traders initiate responses to them and when the executions are attempted; the quotes are simply not reliable. In fact, in many markets, over 95% of quotes are canceled rather than executed. Devise and discuss some strategies for making quotes more reliable to traders.

6. In Section 13.7, we discussed the need for trading firms to question and more closely examine "traders working unusually long hours and working into the late night," and to ensure that traders take "vacations extending at least two consecutive weeks during which traders are denied all access to trading, reporting and control systems are useful for detecting unauthorized trading." Why are such policies helpful to limit rogue trading?
7. Many observers of the market, including Michael Lewis, author of the book Flash Boys, have accused HFT firms of front-running by observing order flow to the market and using those orders and their superior technology to place and execute their own orders before earlier "slow orders" are executed. At least for the scenario described here, does the accusation of HFT front-running seem justified? Why or why not?
8. How might retail clients of brokers benefit from payment for order flow?

Mathematics Appendix

A.1 A BRIEF OVERVIEW OF ELEMENTARY STATISTICS

Mean, Variance, and Standard Deviation

The purpose of this appendix is to review several important elementary statistics concepts. To begin with, suppose that we wish to describe or summarize the characteristics or distribution of a single population of values (or sample drawn from a population). Two important characteristics include central location (measured by average, mean, median, expected value, or mode) and dispersion (measured by range, variance, or standard deviation).

In many instances, we will be most interested in the typical value (if it exists) drawn from a population or sample; that is, we are interested in the "location" of the data set. Mean (often referred to as average) or expected values (sometimes referred to as weighted average) are frequently used as measures of location (or central tendency) because they account for all relevant data points and the frequency with which they occur. The arithmetic mean value of a population μ is computed by adding the values x_i associated with each observation i and dividing the result by the number of observations n in the population:

$$\mu = \sum_{i=1}^{n} x_i \div n \tag{A.1}$$

Variance is a measure of the dispersion (variability and sometimes volatility or uncertainty) of values within a data set. In a finance setting, variance is also used as an indicator of risk. Variance is defined as the mean of squared deviations of actual data points from the mean or expected value of a data set. Deviations are squared to ensure that negative deviations do not cancel positive deviations, resulting in zero variances. High variances imply high dispersion of data. This indicates that certain or perhaps many data points are significantly different from mean or expected values. Population and sample variances are computed as follows:

$$\sigma^2 = \sum_{i=1}^{n} (x_i - \mu)^2 \div n \tag{A.2}$$

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n-1} \tag{A.3}$$

Standard deviation is simply the square root of variance. It is also used as a measure of dispersion, risk, or uncertainty. Standard deviation is sometimes easier to interpret than variance because its value is expressed in terms of the same units as the data points themselves rather than their squared values. High standard deviations and high variances imply high dispersion of data. Standard deviations for populations and samples are computed as follows:

$$\sigma = \sqrt{\sum_{i=1}^{n} (x_i - \mu)^2 \div n} \tag{A.4}$$

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n-1}} \tag{A.5}$$

Co-movement Statistics

A joint probability distribution is concerned with probabilities associated with each possible combination of outcomes drawn from two sets of data. Covariance measures the mutual variability of outcomes selected from each set; that is, covariance measures the relationship between variability in one data set relative to variability in the second data set, where variables are selected one at a time from each data set and paired. If large values in one data set seem to be associated with large values in the second data set, covariance is positive; if large values in the first data set seem to be associated with small values in the second data set, covariance is negative. If data sets are unrelated, covariance is zero. Covariance between data set x and data set y may be measured as follows, depending on whether one is interested in covariance of a population or a sample or expected covariance:

$$\sigma_{x,y} = \sum_{i=1}^{n} (x_i - \mu_x)(y_i - \mu_y) \div n \tag{A.6}$$

$$s_{x,y} = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{n-1} \tag{A.7}$$

The sign associated with covariance indicates whether the relationship associated with the data in the sets are direct (positive sign), inverse (negative sign), or independent (covariance is zero). The absolute value of covariance measures the strength of the relationship between the two data sets. However, the absolute value of covariance can often be more easily interpreted when it is expressed relative to the standard deviations of each of the two data sets. That is, when we divide covariance by the product of the standard

deviations of each of the data sets, we obtain the sample coefficient of correlation $\rho_{x,y}$ as follows:

$$\rho_{x,y} = \frac{\sigma_{x,y}}{\sigma_x \sigma_y} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})/(n-1)}{\sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}} \cdot \sqrt{\frac{\sum_{i=1}^{n}(y_i - \overline{y})^2}{n-1}}} \tag{A.8}$$

A correlation coefficient equal to 1 indicates that the two data sets are perfectly positively correlated; that is, their changes are always in the same direction, by the same proportions, with 100% consistency. Correlation coefficients will always range between -1 and $+1$. A correlation coefficient of -1 indicates that the two data sets are perfectly inversely correlated; that is, their changes are always in the opposite direction, by the same proportions with 100% consistency. The closer a correlation coefficient is to -1 or $+1$, the stronger is the relationship between the two data sets. A correlation coefficient equal to zero implies independence (no relationship) between the two sets of data.

The correlation coefficient may be squared to obtain the coefficient of determination (also referred to as r^2 or r-square in some statistics texts and here as ρ^2). The coefficient of determination is the proportion of variability in one data set that is explained by or associated with variability in the second data set. For example, ρ^2 equal to 0.35 indicates that 35% of the variability in one data set is explained in a statistical sense by variability in the second data set.

Hypothesis Testing

In this section, we discuss the process of induction to form testable hypotheses or theories from specific observations. These hypotheses or theories are useful if they provide a means to make meaningful predictions. Normally, testing a theory involves the collection of additional observations to determine whether they support the theory's predictions. If the additional observations do not confirm the predictions, then one has grounds for rejecting the theory. The observations collected to test a theory are usually represented by numbers or data. In most cases, statistical inference concerns the generalization of sample results to a population.

In many instances, one might make use of statistical inference to test a hypothesis. By convention, a hypotheses test usually involves formulation of a *null* (or maintained) hypothesis along with a competing *alternative* (research or challenging) hypothesis. The null hypothesis H_0 usually is the claim that the population parameter equals some "maintained" value (note that null frequently implies no difference, no impact or nothing). The null hypothesis normally includes an equality sign or either $\leq$ or $\geq$ signs. The alternative hypothesis H_A is the claim that the population parameter differs from the maintained value. The alternative hypothesis normally includes a strict inequality sign. Such tests are usually structured in a conservative manner such that the burden of proof is on the

alternative hypothesis. One supports the research or alternative hypothesis by demonstrating the null hypothesis to be false (rejecting the null hypothesis). One rejects the null hypothesis only when the probability of its being true is sufficiently low (the conventional probability, known as a level of significance, is 0.05 or 0.01). In some instances, the appropriate level of significance for a hypothesis test can be based on the relative costs of rejecting the null hypothesis when it is true or accepting it when it is false.

One might list the steps of a typical statistical hypothesis test as follows:

1. Define the null hypothesis, H_0.
2. Define the alternative hypothesis, H_A.
3. Determine a level of significance, α, for the test.
4. Determine the decision rule or test statistic along with acceptance or rejection regions or critical value based on α.
5. Perform computations.
6. Form conclusions.

The *decision rule* or *test statistic* is a given function of a measurement drawn from the sample on which the statistical decision will be based. The *rejection region* consists of those values of the test statistic that will lead to rejection of the null hypothesis. The *critical value* marks the boundary between the acceptance and rejection regions.

An experiment involving a given sample drawn from a population has some probability of resulting in an erroneous conclusion. Thus, one's hypothesis test may lead to an incorrect rejection of a true null hypothesis (*Type I error*) or an incorrect retaining of a false null hypothesis (*Type II error*). The *power of a test* refers to the probability of not committing a Type II error. This is equivalent to the probability of accepting the alternative hypothesis when it is correct. A test is considered to be superior when its power is higher.

Statistics are most useful for empirical studies in finance. This chapter provides, at best, a very superficial overview of a few of the applications of statistical methodology in finance. The reader is advised to consult a more comprehensive statistics text for a more detailed presentation of statistical methodology and its applications to financial problems.

Hypothesis Testing: Two Populations

Here, we are concerned with comparing two means, μ_1 and μ_2 for populations 1 and 2 with standard deviations σ_1 and σ_2. We shall assume that our samples are independent and drawn from populations whose data are normally distributed. Our test will be based on samples of sizes n_1 and n_2. The samples will have means and variances equal to $\overline{x}_1$ and $\overline{x}_2$ and $s^2{}_1$ and $s^2{}_2$, respectively. We will base our testing methodologies on test statistics and distributions somewhat different from those used earlier. Suppose that we wanted to test whether the means of two populations were different based on samples drawn from those populations. Our hypotheses and test statistics might be as follows:

$$H_0: \quad \mu_1 = \mu_2$$

$$H_A: \quad \mu \neq \mu_2$$

$$t = \frac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\left[\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 - 2)}\right]\left[\frac{n_1 + n_2}{n_1 \cdot n_2}\right]}} \tag{A.9}$$

where $s^2{}_1$ and $s^2{}_2$ are the sample variances. If we are testing whether $\overline{x}_1$ and $\overline{x}_2$ are equal, then our hypothesized difference in means $\mu_1 - \mu_{2\,=}\, 0$ is used for computing our test statistic. Our test statistic assumes that our data follows a student-t distribution.

A variety of other types of tests involving samples from two populations can be constructed as well. For example, tests can be developed to determine whether variances differ, other tests can be based on samples with matched pairs of observations, and so on. A statistics or econometrics text can be consulted to provide additional testing methodologies.

Introduction to Simple OLS Regression

Regressions are used to determine relationships between a dependent variable and one or more independent variables. A simple regression is concerned with the relationship between a dependent variable and a single independent variable; a multiple regression is concerned with the relationship between a dependent variable and a series of independent variables. A linear regression is used to describe the relationship between the dependent and independent variable(s) to a linear function or line (or hyperplane in the case of a multiple regression).

The simple ordinary least squares regression (simple OLS) takes the following form:

$$y_t = \hat{b}_0 + \hat{b}_1 x_t + \varepsilon_{i,t} \tag{A.10}$$

The OLS regression coefficients $\hat{b}_0$ and $\hat{b}_1$ are derived by minimizing the variance of errors in fitting the curve (or m-dimensional surface for multiple regressions involving m variables). Since the expected value of error terms equals zero, this derivation is identical to minimizing error terms squared. Regression coefficient $\hat{b}_1$ is simply the covariance between y and x divided by the variance of x; $\hat{b}_1$ and $\hat{b}_0$ are found as follows:

$$\hat{b}_1 = \frac{\sigma_{x,y}^2}{\sigma_x^2} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}[(x_i - \overline{x})]^2} \tag{A.11}$$

$$\hat{b}_0 = \overline{y} - \hat{b}_1 \overline{x} \tag{A.12}$$

Appropriate use of the OLS requires the following assumptions:

- Dependent variable values are distributed independently of one another.
- The variance of x is approximately the same over all ranges for x.
- The variance of error term values is approximately the same over all ranges of x.
- The expected value of each disturbance or error term equals zero.

Violations in these assumptions will weaken the validity of the results obtained from the regression and may necessitate either modifications to the OLS regression or different statistical testing techniques.

A simple regression is concerned with the relationship between a dependent variable and a single independent variable. Regression coefficients $\hat{b}_0$ and $\hat{b}_1$ represent the vertical intercept and the slope in the statistical linear relationship between the dependent variable y_i and the independent variable x_i. Thus the vertical intercept $\hat{b}_0$ represents the regression's forecasted value for y_i when x_i equals zero and the slope of the regression $\hat{b}_1$ represents the change in $\hat{y}_i$ (the value forecast by the regression for y_i) induced by a change in x_i. The error term $\hat{\varepsilon}_i$ represents the vertical distance between the value $\hat{y}_i$ forecasted by the regression based on its true value y_i; that is, $\hat{\varepsilon}_i = y_i - \hat{y}_i$. The OLS regression minimizes the sum or average of these error terms squared. The size of the sum of the squared errors (often called SSE or, when divided by $(n - 2)$, the variance of errors σ^2_{ε}) will be used to measure the predictive strength of the regression equation. A regression with smaller error terms or smaller $\sigma^2_{\hat{\varepsilon}}$ is likely to be a better predictor, all else held constant.

Once we have determined the statistical relationship between y_i and x_i based on our OLS, our next problem is to measure the strength of the relationship, or its significance. One of the more useful indicators of the strength of the regression is the coefficient of determination or ρ^2 statistic. The coefficient of determination (often referred to as r-square) represents the proportion of variation of variable y that is explained by its regression on x. It is determined as follows:

$$\rho^2_{x,y} = \frac{(\sigma_{x,y})^2}{\sigma^2_x \sigma^2_y} = \frac{\left[\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})\right]^2}{\left[\sum_{i=1}^{n}(x_i - \overline{x})\right]^2 \left[\sum_{i=1}^{n}(y_i - \overline{y})\right]^2} \tag{A.13}$$

This coefficient of determination may also be expressed as either of the following:

$$\rho^2_{x,y} = \frac{\text{Total Variation in } y \text{ Explained by the Regression}}{\text{Total Variation in } y} \tag{A.14}$$

$$\rho^2_{x,y} = \frac{\sum_{i=1}^{n}(y_i - \overline{y})^2 - \sum_{i=1}^{n}\varepsilon_i^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \tag{A.15}$$

The sum $\sum (y_i - \overline{y})^2$ represents total variation in y; the sum $\sum \hat{\varepsilon}_i^2$ represents the variation in y not explained by the regression on x.

Assume that there exists for a population a true OLS regression equation $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ representing the relationship between y_i and x_i, without measurement or sampling error. However, we propose the regression $y_i = \hat{b}_0 + \hat{b}_1 x_i + \hat{\varepsilon}_i$, whose ability to represent the true relationship between y_i and x_i is a function of our ability to measure and sample properly. Our sampling coefficients $\hat{b}_0$ and $\hat{b}_1$ are merely estimates for the true coefficients β_0 and β_1 and they may vary from sample to sample. It is useful to know the significance of each of these sampling coefficients in explaining the relationship between y_i and x_i.

Our estimate $\hat{b}_1$ for the slope coefficient β_1 might vary from regression to regression, depending on how our sample varies. Our estimates for $\hat{b}_1$ will follow a *t*-distribution if our sample of y_i's is large or normally distributed; if our sample is sufficiently large, our estimates for β_1 may be characterized as normally distributed. One potential test of the significance of our coefficient estimate $\hat{b}_1$ is structured as follows:

$$\begin{aligned} H_0&: \ \beta_1 \leq 0 \\ H_A&: \ \beta_1 > 0 \end{aligned}$$

Our null hypothesis is that y is unrelated or inversely related to x; our alternative hypothesis is that y is directly related to x. The first step in our test is to compute the standard error $se(\hat{b}_1)$ of our estimate for β_1 as follows:

$$se(\hat{b}_1) = \sqrt{\frac{1}{n}\frac{\sigma_{\hat{\varepsilon}}^2}{\sigma_x^2}} = \sqrt{\frac{\frac{\sum_{i=1}^{n}\hat{\varepsilon}_i^2}{n-2}}{\sum_{i=1}^{n}(x_i-\bar{x})^2}} \tag{A.16}$$

The standard error for $\hat{b}_1$ is, in a sense, an indicator of our level of uncertainty regarding our estimate for β_1. The numerator within the radical indicates the variability unexplained by the regression; the denominator indicates total variability. Our next step is to find the test statistic for $\hat{b}_1$. This is analogous to standardizing or finding the normal deviate in our earlier hypothesis tests:

$$t(\hat{b}_1) = \frac{\hat{b}_1}{se(\hat{b}_1)} \tag{A.17}$$

We next compare this test statistic to a critical value from a table representing the *t*-distribution (such as Table A.2 at the end of this mathematics appendix).

The process for determining the statistical significance of the vertical intercept $\hat{b}_0$ is quite similar to that for determining the statistical significance for $\hat{b}_1$. We first designate appropriate hypotheses, such as:

$$\begin{aligned} H_0&: \ \beta_0 = 0 \\ H_A&: \ \beta_0 \neq 0 \end{aligned}$$

The primary difference in the process is in determining $se(\hat{b}_0)$:

$$se(\hat{b}_0) = \sqrt{\frac{\frac{\sum_{i=1}^{n}\hat{\varepsilon}_i^2}{n-2} \cdot \sum_{i=1}^{n}x_i^2}{n \cdot \sum_{i=1}^{n}(x_i-\bar{x})^2}} \tag{A.18}$$

Next, we find our *t*-statistic as follows:

$$t(\hat{b}_0) = \frac{\hat{b}_0}{se(\hat{b}_0)} \tag{A.19}$$

We then compare the *t*-statistic to the appropriate critical value just as we did when testing the significance of the slope coefficient. This particular test involves two tails, since our alternative hypothesis is a strict inequality. Be certain to make appropriate adjustments to the critical value (for example, divide α by 2 for two-tailed tests) when making comparisons.

A.2 ESSENTIALS OF MATRICES AND MATRIX ARITHMETIC

A *matrix* is simply an ordered rectangular array of numbers. A matrix is an entity that enables one to represent a series of numbers as a single object, thereby providing for convenient systematic methods for completing large numbers of repetitive computations. Such objects are essential for the management of large data structures. Rules of matrix arithmetic and other matrix operations are often similar to rules of ordinary arithmetic and other operations, but they are not always identical. In this text, matrices will usually be denoted with bold uppercase letters. When the matrix has only one row or one column, bold lowercase letters will be used for identification. The following are examples of matrices:

$$\mathbf{A} = \begin{bmatrix} 4 & 2 & 6 \\ 3 & 7 & 4 \\ 8 & -5 & 9 \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} 2 & -3 \\ 3/4 & -1/2 \end{bmatrix} \quad \mathbf{c} = \begin{bmatrix} 1 \\ 5 \\ 7 \end{bmatrix} \quad \mathbf{d} = \begin{bmatrix} 3 \\ 5 \end{bmatrix}$$

The dimensions of a matrix are given by the ordered pair $m \times n$, where m is the number of rows and n is the number of columns in the matrix. The matrix is said to be of *order* $m \times n$ where, by convention, the number of rows is listed first. Thus, **A** is 3×3, **B** is 2×2, **c** is 3×1, and **d** is 1×1. Each number or value in a matrix is referred to as an element. The symbol $a_{i,j}$ denotes the element in row i and column j of matrix **A**, $b_{i,j}$ denotes the element in row i and column j of matrix **B**, and so on. Thus, $a_{3,2}$ is -5 and $c_{2,1} = 5$.

There are specific terms denoting various types of matrices. Each of these particular types of matrices has useful applications and unique properties for working with. For example, a *vector* is a matrix with only one row or one column. Thus, the dimensions of a vector are $1 \times n$ or $m \times 1$. Matrix **c** above is a column vector, of order 3×1. A $1 \times n$ matrix is a row vector with n elements. The column vector has one column and the row vector has one row. A *scalar* is a matrix with exactly one element. Matrix **d** is a scalar. A *square matrix* has the same number of rows and columns ($m = n$). Matrix **A** is square and of order 2. The set of elements extending from the upper left-most corner to the lower right-most corner in a square matrix are said to be in the *principal diagonal*. For each of these elements $i_{i,j}$, $i = j$. Principal diagonal elements of square matrix **A** are $a_{1,1} = 4$, $a_{2,2} = 7$, and $a_{3,3} = 9$. Matrices **B** and **d** are also square matrices.

A *symmetric matrix* is a square matrix where $c_{i,j}$ equals $c_{j,i}$ for all i and j; that is, the ith element in each row equals the jth element in each column. Scalar **d** and matrices **H**, **I**, and **J** below are all symmetric matrices. A *diagonal matrix* is a symmetric matrix whose elements off the principal diagonal are zero, where the *principal diagonal* contains the series of

elements where $i = j$. Scalar **d** and matrices **H** and **I** in the following are all diagonal matrices. An *identity* or *unit* matrix is a diagonal matrix consisting of ones along the principal diagonal. Both matrices **H** and **I** following are diagonal matrices; **I** is the 3×3 identity matrix:

$$\mathbf{H} = \begin{bmatrix} 13 & 0 & 0 \\ 0 & 11 & 0 \\ 0 & 0 & 10 \end{bmatrix} \quad \mathbf{I} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \mathbf{J} = \begin{bmatrix} 1 & 7 & 2 \\ 7 & 5 & 0 \\ 2 & 0 & 4 \end{bmatrix}$$

Matrix Arithmetic

Matrix arithmetic provides for standard rules of operation just as conventional arithmetic. Matrices can be added or subtracted if their dimensions are identical. Matrices **A** and **B** add to **C** if $a_{i,j} + b_{i,j} = c_{i,j}$ for all i and j:

$$\underset{\mathbf{A}}{\begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{bmatrix}} + \underset{\mathbf{B}}{\begin{bmatrix} b_{1,1} & b_{1,2} & \cdots & b_{1,n} \\ b_{2,1} & b_{2,2} & \cdots & b_{2,n} \\ \vdots & \vdots & \vdots & \vdots \\ b_{m,1} & b_{m,2} & \cdots & b_{m,n} \end{bmatrix}} = \underset{\mathbf{C}}{\begin{bmatrix} c_{1,1} & c_{1,2} & \cdots & c_{1,n} \\ c_{2,1} & c_{2,2} & \cdots & c_{2,n} \\ \vdots & \vdots & \vdots & \vdots \\ c_{m,1} & c_{m,2} & \cdots & c_{m,n} \end{bmatrix}}$$

For example:

$$\underset{\mathbf{A}}{\begin{bmatrix} 2 & 4 & 9 \\ 6 & 4 & 25 \\ 0 & 2 & 11 \end{bmatrix}} + \underset{\mathbf{B}}{\begin{bmatrix} 3 & 0 & 6 \\ 2 & 1 & 3 \\ 7 & 0 & 4 \end{bmatrix}} = \underset{\mathbf{C}}{\begin{bmatrix} 5 & 4 & 15 \\ 8 & 5 & 28 \\ 7 & 2 & 15 \end{bmatrix}}$$

Note that each of the three matrices is of dimension 3×3 and that each of the elements in matrix **C** is the sum of corresponding elements in matrices **A** and **B**. The process of subtracting matrices is similar, where $d_{i,j} - e_{i,j} = f_{i,j}$ for $\mathbf{D} - \mathbf{E} = \mathbf{F}$:

$$\underset{\mathbf{D}}{\begin{bmatrix} d_{1,1} & d_{1,2} & \ldots & d_{1,n} \\ d_{2,1} & d_{2,2} & \ldots & d_{2,n} \\ \vdots & \vdots & \vdots & \vdots \\ d_{m,1} & d_{m,2} & \ldots & d_{m,n} \end{bmatrix}} - \underset{\mathbf{E}}{\begin{bmatrix} e_{1,1} & e_{1,2} & \ldots & e_{1,n} \\ e_{2,1} & e_{2,2} & \ldots & e_{2,n} \\ \vdots & \vdots & \vdots & \vdots \\ e_{m,1} & e_{m,2} & \ldots & e_{m,n} \end{bmatrix}} = \underset{\mathbf{F}}{\begin{bmatrix} f_{1,1} & f_{1,2} & \ldots & f_{1,n} \\ f_{2,1} & f_{2,2} & \ldots & f_{2,n} \\ \vdots & \vdots & \vdots & \vdots \\ f_{m,1} & f_{m,2} & \ldots & f_{m,n} \end{bmatrix}}$$

For example:

$$\underset{\mathbf{D}}{\begin{bmatrix} 9 & 4 & 9 \\ 6 & 4 & 8 \\ 5 & 2 & 9 \end{bmatrix}} - \underset{\mathbf{E}}{\begin{bmatrix} 5 & 0 & 6 \\ 2 & 1 & 6 \\ 5 & 0 & 9 \end{bmatrix}} = \underset{\mathbf{F}}{\begin{bmatrix} 4 & 4 & 3 \\ 4 & 3 & 2 \\ 0 & 2 & 0 \end{bmatrix}}$$

Now consider a third matrix operation. The *transpose* $\mathbf{A}^T$ of matrix $\mathbf{A}$ is obtained by interchanging the rows and columns of matrix $\mathbf{A}$. Each $a_{i,j}$ becomes $a_{j,i}$. The following represent matrix $\mathbf{A}$ and its transpose $\mathbf{A}^T$:

$$\underset{\mathbf{A}}{\begin{bmatrix} 1 & 8 & 9 \\ 6 & 4 & 25 \\ 3 & 2 & 35 \end{bmatrix}} \quad \underset{\mathbf{A}^T}{\begin{bmatrix} 1 & 6 & 3 \\ 8 & 4 & 2 \\ 9 & 25 & 35 \end{bmatrix}}$$

The transpose of a column vector is a row vector:

$$\underset{\mathbf{y}}{\begin{bmatrix} 9 \\ 6 \\ 3 \\ 7 \end{bmatrix}} \quad \underset{\mathbf{y}^T}{\begin{bmatrix} 9 & 6 & 3 & 7 \end{bmatrix}}$$

Similarly, the transpose of a row vector is a column vector. Note that the transpose $\mathbf{V}^T$ of a symmetric matrix $\mathbf{V}$ is $\mathbf{V}$:

$$\mathbf{V} = \begin{bmatrix} 0.09 & 0.01 & 0.04 \\ 0.01 & 0.16 & 0.10 \\ 0.04 & 0.10 & 0.64 \end{bmatrix} \quad \mathbf{V}^T = \begin{bmatrix} 0.09 & 0.01 & 0.04 \\ 0.01 & 0.16 & 0.10 \\ 0.04 & 0.10 & 0.64 \end{bmatrix} = \mathbf{V}$$

Multiplication of Matrices

Two matrices $\mathbf{A}$ and $\mathbf{B}$ can be multiplied to obtain the product $\mathbf{AB} = \mathbf{C}$ if the number of columns in the first matrix $\mathbf{A}$ equals the number of rows $\mathbf{B}$ in the second.[1] If matrix $\mathbf{A}$ is of dimension $m \times n$ and matrix $\mathbf{B}$ is of dimension $n \times q$, the dimensions of the product matrix $\mathbf{C}$ will be $m \times q$. Each element $c_{i,k}$ of matrix $\mathbf{C}$ is determined by the following sum:

$$c_{i,k} = \sum_{j=1}^{n} a_{i,j} b_{j,k}$$

For example, consider the following product:

$$\underset{\mathbf{A}}{\begin{bmatrix} 7 & 4 & 9 \\ 6 & 4 & 12 \\ 3 & 2 & 17 \end{bmatrix}} \quad \underset{\mathbf{B}}{\begin{bmatrix} 7 & 6 \\ 5 & 1 \\ 9 & 12 \end{bmatrix}} = \underset{\mathbf{C}}{\begin{bmatrix} 150 & 154 \\ 170 & 184 \\ 184 & 224 \end{bmatrix}}$$

[1] If it is possible to multiply two matrices, they are said to be conformable for multiplication. Any matrix can be multiplied by a scalar, where the product is simply each element times the value of the scalar.

Matrix **C** in the above is found as follows:

$$\underset{\mathbf{A}}{\begin{bmatrix} 7 & 4 & 9 \\ 6 & 4 & 12 \\ 3 & 2 & 17 \end{bmatrix}} \quad \underset{\mathbf{B}}{\begin{bmatrix} 7 & 6 \\ 5 & 1 \\ 9 & 12 \end{bmatrix}} = \underset{\mathbf{C}}{\begin{bmatrix} (7 \cdot 7) + (4 \cdot 5) + (9 \cdot 9) & (7 \cdot 6) + (4 \cdot 1) + (9 \cdot 12) \\ (6 \cdot 7) + (4 \cdot 5) + (12 \cdot 9) & (6 \cdot 6) + (4 \cdot 1) + (12 \cdot 12) \\ (3 \cdot 7) + (2 \cdot 5) + (17 \cdot 9) & (3 \cdot 6) + (2 \cdot 1) + (17 \cdot 12) \end{bmatrix}}$$

Note that the number of columns (3) in matrix **A** equals the number of rows in matrix **B**. Also note that the number of rows in matrix **C** equals the number of rows in matrix **A**; the number of columns in **C** equals the number of columns in matrix **B**.

Inverting Matrices

An *inverse* matrix $\mathbf{A}^{-1}$ exists for the square matrix **A** if the product $\mathbf{A}^{-1}\mathbf{A}$ or $\mathbf{A}\mathbf{A}^{-1}$ equals the identity matrix **I**. Consider the following product:

$$\underset{\mathbf{A}}{\begin{bmatrix} 2 & 4 \\ 8 & 1 \end{bmatrix}} \quad \underset{\mathbf{A}^{-1}}{\begin{bmatrix} \frac{-1}{30} & \frac{2}{15} \\ \frac{4}{15} & \frac{-1}{15} \end{bmatrix}} = \underset{\mathbf{I}}{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}$$

One means for finding the inverse matrix $\mathbf{A}^{-1}$ for matrix **A** is through the use of a process called the *Gauss-Jordan Method*. This method will be performed on matrix **A** by first augmenting it with the identity matrix as follows:

$$\left[\begin{array}{cc:cc} 2 & 4 & 1 & 0 \\ 8 & 1 & 0 & 1 \end{array}\right] \tag{A}$$

For the sake of convenience, call the above augmented matrix **A** temporarily. Now, a series of row operations (addition, subtraction, or multiplication of each element in a row) will be performed such that the identity matrix replaces the original matrix **A** (on the left side). The right-side elements will comprise the inverse matrix $\mathbf{A}^{-1}$. Thus, in our final augmented matrix, we will have ones along the principal diagonal on the left side and zeros elsewhere; the right side of the matrix will comprise the inverse of **A**. Allowable row operations include the following:

1. Multiply a given row by any constant. Each element in the row must be multiplied by the same constant.
2. Add a given row to any other row in the matrix. Each element in a row is added to the corresponding element in the same column of another row.
3. Subtract a given row from any other row in the matrix. Each element in a row is subtracted from the corresponding element in the same column of another row.
4. Any combination of the above. For example, a row may be multiplied by a constant before it is subtracted from another row.

Our first row operation will serve to replace the upper left corner value with a one. We multiply row 1 in **A** (row 1*A*) by 0.5 to obtain the following:

$$\left[\begin{array}{cc:cc} 1 & 2 & 0.5 & 0 \\ 8 & 1 & 0 & 1 \end{array}\right] \begin{array}{l} 1A \cdot 0.5 = 1B \\ 2A \end{array}$$

where row 1*B* replaces row 1*A*. Now we obtain a zero in the lower left corner by multiplying row 2 in **A** by 1/8 and subtracting the result from our new row 1 to obtain matrix **B** as follows:

$$\left[\begin{array}{cc:cc} 1 & 2 & 0.5 & 0 \\ 0 & \frac{15}{8} & 0.5 & \frac{-1}{8} \end{array}\right] \begin{array}{l} 1A \cdot 0.5 = 1B \\ 1B - (2A \cdot \frac{1}{8}) = 2B \end{array} \tag{B}$$

Next, we obtain a 1 in the lower right corner of the left side of the matrix by multiplying row 2*B* by 8/15:

$$\left[\begin{array}{cc:cc} 1 & 2 & 0.5 & 0 \\ 0 & 1 & \frac{4}{15} & \frac{-1}{15} \end{array}\right] \begin{array}{l} 1A \cdot 0.5 = 1B \\ 2B \cdot \frac{8}{15} = 2C \end{array}$$

We obtain a zero in the upper right corner of the left side matrix by multiplying row 2 above by 2 and subtracting from row 1 in **B**:

$$\left[\begin{array}{cc:cc} 1 & 0 & \frac{-1}{30} & \frac{2}{15} \\ 0 & 1 & \frac{4}{15} & \frac{-1}{15} \end{array}\right] \begin{array}{l} 1B - (2C \cdot 2) = 1C \\ 2B \cdot \frac{8}{15} = 2C \end{array} \tag{C}$$

The left side of augmented matrix **C** is the identity matrix; the right side of **C** is $\mathbf{A}^{-1}$. Because matrices cannot be divided as numbers are in arithmetic, one performs an analogous operation by inverting the matrix intended to be the "divisor" and postmultiplying this inverse by the first matrix to obtain a quotient. Thus, instead of dividing **A** by **B** to obtain **D**, one inverts **B** and obtains **D** by the product $\mathbf{AB}^{-1} = \mathbf{D}$. This concept is extremely useful for many types of algebraic manipulations.

Solving Systems of Equations

Matrices can be very useful in arranging systems of equations. Consider, for example, the following system of equations:

$$0.05x_1 + 0.12x_2 = 0.05$$

$$0.10x_1 + 0.30x_2 = 0.08$$

This system of equations can be represented as follows:

$$\underset{\mathbf{C}}{\begin{bmatrix} 0.05 & 0.12 \\ 0.10 & 0.30 \end{bmatrix}} \quad \underset{\mathbf{x}}{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}} = \underset{\mathbf{s}}{\begin{bmatrix} 0.05 \\ 0.08 \end{bmatrix}}$$

We are not able to divide **C** by s to obtain **x**; instead, we invert **C** to obtain $\mathbf{C}^{-1}$ and multiply it by **s** to obtain **x**:

$$\mathbf{C}^{-1}\mathbf{s} = \mathbf{x}$$

Therefore, to solve for vector **x**, we first invert **C** by augmenting it with the identity matrix:

$$\left[\begin{array}{cc:cc} 0.05 & 0.12 & 1 & 0 \\ 0.10 & 0.30 & 0 & 1 \end{array}\right] \tag{A}$$

$$\left[\begin{array}{cc:cc} 1 & 2.4 & 20 & 0 \\ 0 & 0.6 & -20 & 10 \end{array}\right] \quad \begin{array}{l} RowB1 = A1 \cdot 20 \\ RowB2 = (10 \cdot A2) - B1 \end{array} \tag{B}$$

$$\underset{\mathbf{I}}{\left[\begin{array}{cc:cc} 1 & 0 & 100 & -40 \\ 0 & 1 & \dfrac{-100}{3} & \dfrac{50}{3} \end{array}\right]}_{\mathbf{C}^{-1}} \quad \begin{array}{l} RowC10 = B1 + -(2.4 \cdot C2) \\ RowC2 = B2 \cdot 5/3 \end{array} \tag{C}$$

Thus, we obtain vector **x** with the following product:

$$\underset{\mathbf{C}^{-1}}{\begin{bmatrix} 100 & -40 \\ \dfrac{-100}{3} & \dfrac{50}{3} \end{bmatrix}} \quad \underset{\mathbf{s}}{\begin{bmatrix} 0.05 \\ 0.08 \end{bmatrix}} = \underset{\mathbf{x}}{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}} = \underset{\mathbf{x}}{\begin{bmatrix} 1.8 \\ \dfrac{-1}{3} \end{bmatrix}} \tag{D}$$

Thus, we find that $x_1 = 1.8$ and $x_2 = -1/3$.

A.3 DERIVATIVES OF POLYNOMIALS

The derivative from calculus can be used to determine rates of change or slopes. They are also useful for finding function maxima and minima. For those functions whose slopes are constantly changing, the derivative is an instantaneous rate of change; that is, the change in y induced by the "tiniest" change in x. Assume that y is given as a function of variable x. If x were to increase by a small (infinitesimal—that is, approaching, though not quite, equal to zero) amount h, by how much would y change? This rate of change is given by the derivative of y with respect to x, which is defined as follows:

$$\frac{dy}{dx} = f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

One type of function that appears regularly in finance is the polynomial function. This type of function defines variable y in terms of a coefficient c (or series of coefficients c_j), variable x (or series of variables x_j), and an exponent n (or series of exponents n_j). Strictly speaking, the exponents in a polynomial equation must be non-negative integers; however, the rules that we discuss here still apply when the exponents assume negative or noninteger values. Where there exists one coefficient, one variable, and one exponent, the polynomial function is written as follows:

$$y = c \cdot x^n$$

For example, let $c = 7$ and $n = 4$. Thus, our polynomial is written as follows: $y = 7x^4$. The derivative of y with respect to x is given by the following function:

$$\frac{dy}{dx} = c \cdot n \cdot x^{n-1}$$

Taking the derivative of y with respect to x in our example, we obtain $dy/dx = 7 \times 4 \times x^{4-1} = 28x^3$. Note that this derivative is always positive when $x > 0$; thus the slope of this curve is always positive when $x > 0$. Consider a second polynomial with more than one term (m terms total). In this second case, there will be one variable x, m coefficients (c_j), and m exponents (n_j):

$$y = \sum_{j=1}^{m} c_j \cdot x^{n_j}$$

The derivative of such a function y with respect to x is given by:

$$\frac{dy}{dx} = \sum_{j=1}^{m} c_j \cdot n_j \cdot x^{n_j - 1}$$

That is, simply take the derivative of each term in y with respect to x and sum these derivatives. Consider a second example, a second-order (the largest exponent is 2) polynomial function given by $y = 5x^2 - 3x + 2$. The derivative of this function with respect to x is: $dy/dx = 10x - 3$. This derivative is positive when $x > 0.3$, negative when $x < 0.3$, and zero when $x = 0.3$. Thus, when $dy/dx > 0$, y increases as x increases; when $dy/dx < 0$, y decreases as x increases; and when $dy/dx = 0$, y may be either minimized or maximized. Also note that y is minimized when $x = 0.3$; at this point, $dy/dx = 0$.

As suggested above, derivatives can often be used to find minimum and maximum values of functions. To find the minimum value of y in function $y = 5x^2 - 3x + 2$, we set the first derivative of y with respect to x equal to zero and then solve for x. For our example, the minimum is found as follows:

$$\begin{aligned} 10x - 3 &= 0 \\ 10x &= 3 \\ x &= \frac{3}{10} \end{aligned}$$

In order to ensure that we have found a minimum (rather than a maximum), we check the second derivative. The second derivative is found by taking the derivative of the first derivative. If the second derivative is greater than zero, we have a minimum value for y (the function is concave up). When the second derivative is less than zero, we have a maximum (the function is concave down). If the second derivative is zero, we have neither a minimum nor a maximum. The second derivative in the above example is given by $d^2y/dx^2 = 10$, also written $f''(x) = 10$. Since the second derivative 10 is greater than zero, we have found a minimum value for y. In many cases, more than one "local" minimum or maximum value will exist.

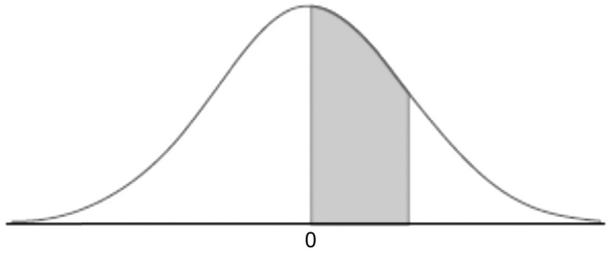
Consider a third example where our second order polynomial is given: $y = -7x^2 + 4x + 5$. The first derivative is $dy/dx = -14x + 4$. Setting the first derivative equal to zero, we find our maximum as follows:

$$\begin{aligned} -14x + 4 &= 0 \\ -14x &= -4 \\ x &= \frac{4}{14} \end{aligned}$$

We check second-order conditions (the second derivative) to ensure that this is a maximum. The second derivative is $d^2y/dx^2 = -14$. Since -14 is less than zero, we have a maximum at $x = 4/14$.

A.4 REFERENCE TABLES

TABLE A.1 Normal Distribution, *z*-values



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0159	0.0199	0.0239	0.0279	0.0319	0.0358
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0909	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2356	0.2389	0.2421	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2703	0.2734	0.2764	0.2793	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3437	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3906	0.3925	0.3943	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.492	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4986	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

TABLE A.2 Student *t*-Distribution

	Right Area, α				
df	**0.100**	**0.050**	**0.025**	**0.010**	**0.005**
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.695	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
∞	1.282	1.645	1.960	2.326	2.576

Examples: The t value for ∞ degrees of freedom that bound a right-tail area of 0.025 is 1.960. The t value for ∞ degrees of freedom that bound left- and right-tail areas (two tails) summing to 0.05 is 1.960.

End-of-Chapter Exercise Solutions

CHAPTER 1

1. Under some, but not all definitions, yes, a futures contract is a security. A futures contract is a publicly traded financial instrument, which qualifies it to be considered a security under some definitions. However, a futures contract might not qualify as a security under the Howey Test. For example, a futures contract normally does not involve an investment of money, even though it does involve deposit of margin. More problematic is that a futures contract does not involve the formation of a common enterprise and the investment into an enterprise.

2. Vigorous competition among professional short-term traders can itself lead to very important benefits for long-term investors, including narrower spreads and greater depth.

3. First, recall the definition of trading: "a security transaction that creates a portfolio position based on an investment decision." While the trade might indeed result from a decision to speculate, it might also result from a long-term investment decision or decisions to hedge or engage in arbitrage. While traders can be speculators who focus on profits derived from price changes, traders can also be arbitrageurs who focus on price discrepancies and hedgers who seek to control risk. Traders might even be executing transactions as an agent on a long-term investor's account, acting as a broker for that investor.

4. You should bid more than $200, say, $1000, because you will be able to obtain the object for $10, the second-highest bid. No one will be willing to bid against you for more than $200 to obtain the object for $1000. Thus, it is important for second-price auction bids to be sealed.

5. **a.** $E[x_{MAX}] = 500{,}000 + (1{,}000{,}000 - 500{,}000) \times 2/(2+1) = 833{,}333$
$E[x_{MIN}] = 500{,}000 + (1{,}000{,}000 - 500{,}000)/(2+1) = 666{,}666$, which is the second-highest bid.

b. $E[x_{MAX}] = 500{,}000 + (1{,}000{,}000 - 500{,}000) \times 3/(3+1) = 875{,}000$
$E[x_{MIN}] = 500{,}000 + (1{,}000{,}000 - 500{,}000)/(3+1) = 625{,}000$
With only three draws from the distribution, the second highest bid is the mean of the highest and the lowest.
$E[x_{2nd\ highest}] = 500{,}000 + (1{,}000{,}000 - 500{,}000) \times 2/(3+1) = 750{,}000$, which is the second-highest bid.

c. $E[x_{MAX}] = 500{,}000 + (1{,}000{,}000 - 500{,}000) \times 999/(999 + 1) = 999{,}500$
$E[x_{MIN}] = 500{,}000 + (1{,}000{,}000 - 500{,}000)/(999 + 1) = 500{,}500$
The second-highest bid expected value is $E[x_{MAX}] - (x_2 - x_1)/(n + 1)$
$E[x_{2nd\ highest}] = 999{,}500 - (1{,}000{,}000 - 500{,}000)/(999 + 1) = 999{,}000$, which is the second-highest bid. Note that the second-highest bid increases as the number of bidders increases.

6. Obviously, the Treasury wants to sell as many bills as possible at the lowest possible prices. The bid-to-cover ratio in this illustration is \$34 billion/\$20 billion = 1.7. First, the \$2.5 billion in noncompetitive bids will be subtracted from the \$20 billion total to determine the stop-out price. Bids will be satisfied from the lowest yield (highest price) until \$17.5 billion in bills have been allocated to the competitive bidders. The stop-out price will be at a yield of 5.30% and all winners (Citigroup, Merrill Lynch, UBS, Deutsche Bank, noncompeting bidders) will pay this same price. Noncompeting bidders will be allocated \$2.5 billion and Deutsche Bank will be allocated \$5.5 billion.

7. a. The cost of placing each bid is the bid's amount. The bidder that pays the most to place a bid receives the auctioned object. For example, the lobbyist that contributes the most has influence and the fans waiting for rock concert tickets receive them. The losers incur the costs of their bids, but receive nothing. In the case of the politician, all contributions are realized as revenue, not just the winning lobbyist's contributions. In the case of the \$1 bill, suppose one bidder opens with a bid for \$0.01. The second counters with a bid for \$0.02. It may seem reasonable for each of them to continue increasing their bids to a maximum of \$1. However, if the losing bid is \$0.99, it may seem reasonable to increase the bid to \$1.01. Any losing bid represents a sunk cost. To avoid the sunk cost, the bid has to be raised. Essentially, such an auction can be thought of as a war of attrition, where all but one bidder eventually realizes that he is better off abandoning his sunk cost and the winning bidder might end up paying far more for the auctioned item than it's worth.

b. Essentially, from the perspective of the bidders, PennyCave offers an all-pay auction, or something quite close to it. Imagine that their auction results in a winner and a set of losers. Suppose that an object, worth \$100 ultimately sells for \$50, with bids starting at \$0 and increasing in \$0.01 increments until it reached a high of \$50. Assume that each bid costs \$1 to place. Although the \$100 object is obtained by the winning bidder for only \$50, the seller (who is also the auctioneer) receives $((50 - 0)/0.01 \times \$1) = \$5000$ in bidding fees, or \$5050 total. If, for example, there were only two bidders, the loser would have paid approximately \$2500 in bidding fees to receive nothing, while the winner would have paid \$2550 for the object worth \$100. Of course, this assumes naïve bidding strategies and no collusion on the parts of all bidders, but the outcome in this scenario could have been much worse, with bidding exceeding the value of the auctioned item as bidders seek to avoid losing sunk bidding costs. Keeping bid increments small (\$0.01) increases the number of bids placed in the auction. The auctioneer expects to earn more revenue from the bid activity than from the winning bid itself; it seeks to maximize the number of auction bids.

8. a. The best bid is 10.03; the best offer is 10.01. Notice that the market at this point is crossed; the best bid exceeds the best offer.

b. Downward sloping as follows: 100 shares at 10.03, 200 shares at 10.02, 300 shares at 10.01, 400 shares at 10.00, 500 shares at 9.99, and 600 shares at 9.98.

c. Upward sloping as follows: 100 shares at 10.01, 200 shares at 10.02, 300 shares at 10.03, 400 shares at 10.04, and 500 shares at 10.05.

d. The market clearing price is 10.02. The demand at this price is 200 shares (100 share bids each at 10.03 and 10.02). The supply at this price is 200 shares (100 share offers each at 10.01 and 10.02). Since supply (200 shares) equals demand (200 shares) for XYZ at the price of 10.02, the market clearing price is 10.02.

e. The 100 share bids each at 10.03 and 10.02 execute at open. The 100 share offers each at 10.01 and 10.02 execute at open.

f. The open NBBO is 10.01−10.03; The NBB is 10.01 and the NBO is 10.03.

9. a. A limit order sets an upper price for a purchase or a lower price limit for a sell, preventing the broker from paying more or accepting less for the security. A stop order instructs the broker to place the buy order once the price has risen above a given level or place the sell order once the price of the security has fallen beneath a given level. The limit order restricts the price; the stop order triggers the order execution.

b. Buy-side market participants are more likely to place market orders. They pick securities to fulfill an investment need, and the market order is more likely to be executed, though perhaps at a less favorable price.

c. Sell-side market participants are more likely to place limit orders. They place quotes in the market in order to supply liquidity and profit from buy-side market participants.

10. This stop-limit order triggers the buy once the price rises to 50, but is executed only if the stock can be purchased for no more than 50.10. Stop orders to buy are often placed when the investor wants to buy the stock on upward price momentum, but the limit is typically placed when the investor wants protection from paying more than she wants for the stock.

11. Batch trading requires that trades accumulate until a specific point in time, or until a specific level of unexecuted trades have accumulated to allow for a market to be called and the orders to be executed. On the other hand, continuous trading allows for trades to execute immediately upon reaching the market. Thus, batch trading slows down market execution, defeating the purpose of high-frequency trading and diluting the technological advantages enjoyed by high-frequency traders.

12. a. The DMM would receive 1,000 shares, the floor broker would receive 600 shares, and the first of the two Electronic Book orders would receive 400 shares, all at a price of 50.00. The second of the two Electronic Book orders would receive no shares.

b. First, the DMM order with setter priority receives 15% of 4,000 or 600 shares for 50.00. Then, the DMM and the floor broker have equal priority on the next 1,000 shares for $50.00. The DMM receives 400 to complete her order for 1,000 shares, and the floor broker receives 600 shares to complete his order. Four hundred shares remain to be allocated between the two Electronic Book orders, with each receiving 200 shares for $50.00.

13. This answer depends on how many shares are being traded and applies only to a prospective purchaser of these shares. With respect to this particular investor's trade, in Market A, the investor will obtain 25,000 shares for $100.00. The final price remains at $50.00. In Market B, the investor will obtain 10,000 shares for $100.00, 10,000 for $100.01, and 5000 for $100.03. The final price in Market B rises to $50.03, more than Market A. Thus, with respect to this order, Market B has less depth than Market A. However, this might not be the case for certain larger orders or even certain smaller orders.

14. There are probably no correct solutions for this rhetorical exercise; it is intended to provoke some opinions.

a. The quick and easy answers are that the mosquito eradication program produces more social benefits and the trading technology program produces more private benefits, particularly if high values are placed on life and on health improvement. However, for the sake of argument, do consider that it is very difficult to measure the social benefits produced by the financial sector (the efforts of Philippon (2015) notwithstanding), partly because much of this value is captured by other industries and individuals. For example, what is the total benefit of a $500 million loan to start a new hospital? Or to build a pharmaceutical factory employing 1000 workers? In fact, to begin with, who among us is even qualified to decide what values to associate with the programs?

b. Financial, especially trading activities tend to be held in low esteem. The general perception is that financial activities do not directly enhance peoples' lives (it is easy to overlook indirect enhancements), are akin to gambling (transferring risk is an important activity in finance), create wealth inequality, and are rife with fraud (finance *is* where the money is, and where it can be stolen).

c. Certainly, many individuals prefer outcomes with high private benefits for themselves to outcomes that produce large social benefits. In fact, this is the nature of the "Prisoner's Dilemma" problem from game theory. A society, however this term might be defined, can mitigate this problem by forming governments to make decisions concerning projects with high social benefits. On the other hand, are the incentives and private benefits consumed by government agents any less harmful than those held and consumed by agents of financial institutions?

CHAPTER 2

1. a. The specialist system was intended to maintain a continuous, orderly, and liquid market for investors. Participants in a trading crowd on the NYSE trading floor are expected to contribute to this effort. Personal contact with each other enables the specialist and floor brokers and traders to better know each other and establish working relationships and trust.

b. The NYSE was an old exchange with well-established practices that worked very well over the years. In many respects, it is more expensive to change a way of doing business to make it more efficient than it is to create a new one without any history. In addition, the exchange was owned by its members, who had invested

their capital and careers into working on the floor of the exchange. It is very difficult to induce members with such heavy investments in this manner of doing business to give up such a large investment.

2. Two reasons. First, as we saw in Chapter 1, the NYSE can be regarded as having characteristics of both order-driven and quote-driven markets, enabling both market makers (with price quotations) and brokers (with market orders) to initiate trading. Second, and perhaps more significantly, the NYSE (in New York City and the Arca Options floor in San Francisco) blends human interaction in floor trading with automated electronic trading. This hybrid model provides for human judgment and assessment, particularly at market opens and closings, the busiest time for the NYSE and when a large proportion of floor trades are executed. Until 2006, but no longer relevant, the NYSE structure blended characteristics of a partnership with characteristics of a corporation, making it a hybrid organization.
3. True: An ATS performs all of the functions of an exchange; that is, an ATS maintains quotes and provides a forum for securities transactions. However, an ATS does not seek to regulate the conduct of traders aside from their use of the trading system, which is the quality that distinguishes it from an exchange.
4. Credit ratings are created to gauge the creditworthiness of the creditor, and more importantly, indicate the likelihood that the relevant debt obligations will be fully satisfied in a timely fashion. If riskier bonds are less liquid, a likely scenario, the credit rating might imply negative information about liquidity risk. A credit rating might be affected by a bond's term to maturity and coupon rate, but the rating would probably be less useful as a gauge of interest rate risk than values for term to maturity and coupon rate.
5. **a.** Like currency, gold is easy to store and transport, divisible, and it is non-perishable. If a one-year selling price for gold can be locked in today with a forward contract, then the relationship between its forward and spot prices should be $F_0 = E[S_1] = S_0 \cdot (1 + i)$ or $S_0 e^{it}$ with continuous compounding. Use the expectations hypothesis for the relationship between spot and forward rates for currency: $F_0 = E[S_1]$. $E[S_1] = \$1100 \cdot 1.03 = \$1133$.
   b. Again, use the expectations hypothesis as above: $F_0 = E[S_1]$. $E[S_1] = \$1100 \cdot 1.03 = \$1133$.
   c. Assuming that the market prices given ($1,100 spot and 3% riskless rate) are determined by supply and demand, and that the assumptions that we noted in part a are true, this expectations hypothesis might provide for a reasonable relationship between the spot and forward prices for gold.
   d. It is important to remember that the forward rate is based on actual market prices of forward contracts whereas an expected spot rate is merely a prediction by an investor, using prevailing information about prices and other factors such as risk preferences and uncertainty to establish. Hence, there are good reasons for the two values to diverge, including inefficiencies in markets for gold or violations in the assumptions that we noted in part a.
6. **a.** Yes; more francs will be required to purchase a dollar: 2.5 rather than 2 now
   b. No; fewer dollars will be required to purchase a franc: 0.4 rather than 0.5 now

c. The percentage revaluation (devaluation) of francs relative to dollars is determined: $\%\text{reval} = [(S_1 - S_0) \div S_0] = [(0.4 - 0.5) \div 0.5] \times 100 = -20\%$

d. The percentage revaluation of the dollar in terms of francs is 25%: $\%\text{reval} = [(2.5 - 2) \div 2] \times 100 = +25\%$

7. a. The value of $1 is equal to the value of £.6. Thus, the value of £1 must be $1/0.6 = $1.6667. Since the value of $1 is ¥108, the value of £1 must be 1.6667 × ¥108 = ¥180. Thus, we could have solved as follows: £1 = ($1 ÷ 0.6) × ¥108 = ¥180.

b. The value of $1 is equal to the value of ¥108. Thus, the value of ¥1 must be $1/108 = $0.0092593. Since the value of £1 is $1.6667, the value of ¥1 must be 1 ÷ (1.6667 × 108) = £0.0055556. Thus, we could have solved as follows: ¥1 = 1 ÷ (1/0.6 × ¥108) = £0.0055556.

c. $1500 = £.6 × 1500 = £900; $1500 = ¥108 × 1500 = ¥162,000.

8. Ignoring market frictions, forward contract prices must equal spot prices for the same underlying asset on the forward contract settlement date to eliminate arbitrage opportunities, which would be a violation of the Law of One Price. If the two prices were to be unequal, an investor could take a long position in the lower-priced contract and an offsetting short position in the higher-priced contract and pocket the price difference. Thus, as time approaches the forward contract settlement date, the spot price will approach the settlement price of the contract.

9. Futures contracts differ from forward contracts in that they are standardized and trade on exchanges, and they typically provide for margin and marking to the market. That is, futures contracts are traded on exchanges such as the Chicago Board of Trade and the New York Mercantile Exchange. In order to ensure liquidity in futures markets, exchanges limit the number that they trade and maintain standardized terms to ensure demand for each of the contracts that they elect to trade. Margin requirements are instituted to ensure that contract participants fulfill their obligations to the exchange, whereas forward market participants rely on good faith, reputation, and the court system. Generally, a futures market participant must vary her margin, depending on the trading environment and the prices of the underlying assets or rates. This variance of margin, which increases as the value of the investor's obligation increases and decreases as the obligation decreases, is referred to as marking to the market.

10. Technically, the CEO did not sell his shares, though he divested himself of the return risk associated with share ownership. By engaging this equity swap, the CEO reduced his risk in his employing company without having to report a sale of shares (although Weil did voluntarily report this transaction, and the SEC now does require such reporting). This meant that the CEO was not subject to capital gains taxes at the time of the transaction (however, this tax benefit has since been eliminated by the IRS). The CEO is not likely to bear the selling price consequences associated with an insider sell transaction. Furthermore, the CEO maintained his level of voting control in the company's shares because he still owned them. Thus, in a sense, the equity swap permitted the CEO the opportunity to, in effect, execute a sale of shares without bearing most of the undesirable consequences associated with the sale.

11. The following set of transactions, with the first three in any order, will produce an arbitrage profit:

Borrow CHF1000 now in Switzerland at 12%; repay at Time One
Buy USD625 at $S_0 = 1.6$ now for CHF1000

Loan USD625 at 10%; Collect proceeds at Time One
Sell USD622.22 at Time One at $F_1 = 1.8$ for CHF1120

The time zero CHF amount can be any amount, though either the time zero USD amount must be 0.625 times the CHF amount or the time 1 USD amount must be 0.5556 times the CHF amount. Regardless, the investor borrows CHF, converts to USD, and lends USD. The time one amounts are determined by the forward and interest rates. This exercise the concept, not discussed in the chapter, called Interest Rate Parity, which characterizes the arbitrage-free relationship among spot rates, forward rates, and interest rates.

CHAPTER 3

1. First, brokerage firms regularly use such promotions to attract new permanent customers, which then might pay higher stock-trading commissions after the promotions end. Second, we discussed payment for order flow in the chapter, a major source of revenue for many brokerage firms such as Ameritrade. Retail customer orders are especially desirable because they are assumed to be uninformed, creating no adverse selection risk for the dealer. Accordingly, retail brokers route nearly all of their customer market orders to internalizers pursuant to Payment for Order Flow arrangements. Some brokerage firms impose annual account fees on customers, or high "back-end" fees when customers close down accounts. Brokerage firms might use low commissions to "hook" clients with the hope that they can sell them premium services such as wealth management services. Many customers maintain high cash balances, on which brokerage firms can earn significant interest for themselves. Charles Schwab has earned as much as 45% of its revenues from this source. Some firms will charge much higher commissions for other products such as mutual funds and bonds. In addition, any transaction that places securities in the client's account held in street name by the broker produces short-sell opportunities for the broker's other clients, thereby producing lending and interest-bearing opportunities for the broker.
2. Fragmentation is the condition in which multiple market venues exist; it is the opposite of a concentrated market. Fragmented markets mean that traders need to search multiple trading venues in their efforts to obtain the best possible bids, offers, and prices. Multiple trading venues can diminish the market's ability to convey useful price-related information to market participants. SIPs collect, compile, and disseminate quotes and prices from all market venues to all interested traders, better enabling them to execute the best possible trades.
3. Volume and the number of stocks reported on by the SIP are obviously far larger than they were during the ticker tape era. More importantly, in the age of the national market system, the SIP has to collect, aggregate, and compile all price and volume information from every exchange and other trading venue in which every covered stock trades. The SIP also has to provide data on the NBBO and quoted spreads.
4. The NBBO of the Bats-traded EFT (exchange traded fund) PAVE is reported on Tape B because Bats is an exchange other than the NYSE. All exchange listed shares other than those listed on the NYSE and NASDAQ are reported on Tape B. As of 2022, Bats is the largest market for ETFs, followed by Arca.

5. Real-time processing allows for transactions to be processed as soon as they are executed, enabling them to produce funds and financial instruments that can be spent, transferred, or otherwise used immediately. Netting is a form of batch processing that requires transactions to accumulate for a period of time, and with the clearing firm or house counterparty to each transaction, clear in a single batch at the end of the period. This process is much less expensive than real-time clearing because a large number of transactions can be cleared simultaneously.
6. Novation occurs when the clearing house assumes the settlement obligations of both counterparties to a transaction, in effect becoming the counterparty to both sides of the transaction. Netting is the simplification process used by clearing institutions of adding all of a given firm's purchases of each security, adding the sales of each security, deducting sells from buys to determine the net change in holdings of that security for the firm, then finally computing the net cash flows associated with all transactions. The net reflects what each clearing member receives or pays.
7. The central counterparty is the counterparty for all trades in a given venue. Thus, individual trader credit risk is irrelevant because only the credit risk of the central counterparty and all of its backers, usually all members, is relevant.
8. For all parts of this problem, we will aggregate trades for clients and then for members of the DTC for the trading day.
   a. Merrill: +100 shares of General; Morgan: –200 shares of General; F-Trade: +100 shares of DotCom and –100 shares of R-Train; Schwalb: –100 shares of DotCom, –100 shares of R-Train, and +100 shares of General
   b. A: +100 shares of General through Merrill and –200 shares of General through Morgan; B: +100 shares of DotCom through F-Trade; C: +100 shares of R-Train through F-Trade and –100 shares of DotCom through Schwalb; D: –100 shares of R-Train through Schwalb and +100 shares of General through Schwalb
   c. City: $5,000 paid to the DTC; Chase: $19,000 received from the DTC; BoB: $14,000 paid to the DTC
   d. 0; Always zero to the central clearing party (or NSCC or DTC) on each end-of-day settlement; the net amounts paid out equal the net amounts received.
9. Real-time clearing would require that every transaction be prefunded in real time, in advance of transactions themselves, and would mean that each institution would need to settle for each transaction in real time.
   a. The benefits to netting would be largely eliminated with real-time clearing because every transaction would need to be cleared independently of all other transactions.
   b. It is likely that the number of out-trades and DKs would increase as the netting-down process is watered down, increasing the numbers of reports that need to be delivered to every reporter on a daily basis. Real-time clearing is likely to result in more clearing and reporting errors.
10. Robinhood reasoned that by reducing its highly speculative and volatile client trading, it would be able to convince the DTCC to reduce its collateral requirements, thereby reducing pressure on the brokerage to raise enormous levels of cash. Alternatively, Robinhood could have approached its investors and lenders for cash infusions, which Robinhood did as well.

11. No. Settlement risk or counterparty risk to the clearinghouse is more a function of the other counterparty than of the instrument or transaction being settled. If the buyer or seller of the instrument defaults by failing to deliver the appropriate securities, the clearinghouse still needs to cover the defaulted position, which might be quite volatile or costly, especially in a market with liquidity issues or volatile interest rates.

CHAPTER 4

1. First, the order can be parsed/sliced and distributed a number of ways to a number of markets. Slices of the order can be submitted to exchange or OTC markets, and broken into small quantities during the day. Algorithms can be used to efficiently slice the order and to minimize the transparency and price impact of the orders. Pieces of the order can be shopped or submitted through other brokers and through other markets, including crossing networks and other block trading facilities. Finally, the dealer can use its own capital to take up any unsold shares by the end of the day.

2. The quoted spread was $4.27 - 4.21 = 0.06$, the half spread was $(4.27 - 4.21)/2 = 0.03$, and the effective half-spread was $|(4.22 - 4.24)| = 0.02$. The realized half-spread, based on the immediately subsequent transaction, $|4.23 - 4.24|$ was 0.01.

3. a. First, total the transactions costs B and set the derivative of total transaction costs with respect to the number of slices equal to zero and solve for X:

$$B = nF + nv\left(\frac{X}{n}\right)^m = 10n + 0.0001\frac{500{,}000^2}{n} = 10n + 25{,}000{,}000n^{-1}$$

$$\frac{dB}{dn} = \frac{d\left[nF + nv\left(\frac{X}{n}\right)^m\right]}{dn} = F - (m-1)v\frac{X^m}{n^m} = 10 - (2-1)(0.0001)\frac{500{,}000^2}{n^2} = 10 - \frac{25{,}000{,}000}{n^2} = 0$$

$$n = \sqrt[m]{\frac{(m-1)vX^m}{F}} = \sqrt{\frac{(0.0001)\cdot 500{,}000^2}{10}} = 1581$$

Thus, we find that the optimal number of slices or transactions is $n = 1581$, with X/n, the size of each transaction equal to 500,000/1581 = 316. One note here: First, noninteger numbers of transactions are not really possible; hence, we have rounded our figures.

b. Total slippage costs equal $nv(X/n)^m = 0.1581 \times 316^2 = \$15{,}811$. Note that fixed administrative transactions costs total to $\$15{,}811 = \$1581 \times 10$ (again, figures are rounded). Total trading costs are $31,622.

c. First, we write the total slippage cost function as $n \times B/n$ as follows:

$$B = nrn^s + nv\left(\frac{X}{n}\right)^m = rn^{s+1} + nv\left(\frac{X}{n}\right)^m = 0.1n^{1.5}$$

$$+0.0001\frac{500{,}000^2}{n} = 0.1n^{1.5} + 25{,}000{,}000n^{-1}$$

Next, we solve for n, the optimal number of order slices to execute as follows:

$$\frac{dB}{dn} = \frac{d\left[nrn^s + nv\left(\frac{X}{n}\right)^m\right]}{dn} = r(s+1)n^s - (m-1)v\frac{X^m}{n^m} = 0.15n^{.5} - \frac{25,000,000}{n^2} = 0$$

Unfortunately, solving for n requires substitution. We substitute and iterate to find that $n = 1944$. Since $X = 500{,}000$, the optimal order size is $500{,}000/1944 = 257$ (allow for rounding to obtain integer values).

4. a. Beagle can observe slice sizes only. Based on its market impact model, these individual execution sizes $c = 214 = X/n$ do not reveal either X or n. We need to find n to solve for X. Beagle cannot observe the number of orders n until the purchase program is complete; therefore it cannot infer total or latent demand for the stock. Optimal slice or execution size will not vary with the size of the total demand with the given slippage cost function.

b. Based on its own revised market impact function, Beagle will assume that the total slippage for the stock purchase program is determined from the following:

$$B = nrn^s + nv\left(\frac{X}{n}\right)^m = rn^{s+1} + nv\left(\frac{X}{n}\right)^m = 0.1n^{1.5} + 0.0001\frac{X^2}{n} = 0.1n^{1.5} + 0.0001X^2n^{-1}$$

Total slippage is minimized when the derivative of the market impact function B is minimized with respect to n. Thus, we will find and set equal to zero the derivative of B with respect to n:

$$\frac{dB}{dn} = \frac{d\left[nrn^s + nv\left(\frac{X}{n}\right)^m\right]}{dn} = r(s+1)n^s - (m-1)v\frac{X^m}{n^m} = 0.15n^{.5} + 0.0001 \cdot \frac{X^2}{n^2} = 0$$

Since slices or orders of size 214 shares were observed by Beagle, who will assume that these slices are of optimal size, $c = X/n = 214$. Now, we can solve for n in the following to obtain the optimal number of orders in this purchase program:

$$0.15n^{0.5} - 0.0001 \times 214^2 = 0$$

$$n = \left(\frac{\left(\frac{X}{n}\right)^m \cdot (1-m)v}{r \cdot (1+s)}\right)^{1/s} = \left(\frac{0.0001 \times (214)^2}{0.15}\right)^2 = 932$$

Answers might vary from 920 to 940 based on rounding to integer values. Thus, 932 slices with 214 shares each indicates a total or latent demand for $932 \times 214 = 199{,}448$ shares, again, subject to rounding errors.

5. a. Payment for order flow creates a conflict of interest between a broker and its client in which the broker directs the client's order to a particular market or wholesaler and receives a payment for having done so. This conflict of interest weakens investor confidence in the market, and, in the extreme and in the long term, causes investors to withdraw their capital and liquidity provision in that market.

b. Payment for order flow, especially as the result of maker-taker pricing policies enhances liquidity by encouraging market makers to serve as counterparties to investor transactions.

6. Arnuk and Saluzzi (2009) note that it "appears that 'fairness' and the equalization of market data speed among co-located firms is an important 'must' for the exchanges, but not so when it comes to all other institutional and retail investors."[1] The implication here is that exchanges providing co-location facilities and equal amounts of connecting cable to all co-location participants reflects some degree of hypocrisy. Does this practice mean that exchanges know that they are providing, for a fee, opportunities for their tenants to obtain advance quote and trading information?

7. a. Substantial empirical evidence does suggest that HFT, in general, increases market liquidity (e.g., Hasbrouck & Saar, 2013), bringing into the market pools of securities when other traders need trade counterparties; reduces trading costs (Harris, 2013); enhances price discovery (Conrad, Wahal, & Xiang, 2015); and reduces trading spreads (Brogaard, Hendershott, & Riordan, 2014). Similarly, Jones (2013) summarizes that HFT in competing markets does "improve market quality, reduce[s] trading costs, and make[s] stock prices more efficient."

b. The following would seem to weaken the argument that HFT is beneficial to securities markets: Kirilenko, Kyle, Samadi, and Tuzun (2017) found that during the extreme volatility period known as the "Flash Crash" in May 2010, HFTs did draw down liquidity as the crash worsened. Most quotes can be withdrawn nearly instantly by HFT traders, which renders many quotes rather meaningless if they cannot be taken by other investors. Over 95% of quotes in stock markets are ultimately canceled. HFT might produce high-volume, "trigger-happy" trading, perhaps fed by algorithmic, technological, or trading errors that could trigger massive one-sided order flows leading to failures, contagion, and market meltdown. There is evidence that some market liquidity and market spreads are illusory (Eggington, Van Ness, & Van Ness, 2016; Coffee, 2014), in large part because of flash trading and quote stuffing, practices often associated with HFT shops. Brogaard, Hendershott, and Riordan (2014) argue that many HFT traders free ride on the costly efforts of analysts and investors by observing and anticipating the order flow of informed investors and then using their superior technology to advance ahead of the informed trader order flow. Eventually, the inability of slower informed traders to compete might be expected to reduce their incentives to trade, thereby reducing their willingness to conduct the valuable research that is required for market efficiency and optimal capital allocation. HFT is a drain on trader resources. The largely socially wasteful "technological arms race" (Budish, Cramton, & Shim, 2015) associated with HFT imposes a substantial cost on the HFT trader that seeks the fastest response to new information. In addition, it imposes an externality on the rest of the industry as each HFT competitor makes similarly large investments to win the race, resulting in a suboptimally large industry investment in technologies ultimately rendered obsolete. For example, the Tabb Group estimated that trading firms spent approximately $1.5 billion in 2013

[1]Arnuk, S., & Saluzzi, J. (2009, December 4). Latency arbitrage: The real power behind predatory high frequency trading. Themis Trading LLC White Paper.

alone to increase trading speeds (Patterson, 2014). Although HFT has enjoyed a profitable run, numerous sources (e.g., McNamara, 2016) observe that a number of HFT shops are experiencing difficulties and closing as profitability drops and risks remain high. Of course, there are arguments that suggest that HFT is unfair to retail investors who lack the resources to compete with institutional investors with access to HFT technology. Perceived unfairness in securities markets can reduce participation, driving out the capital of investors who believe that they are disadvantaged.

8. A speed bump, a device placed by an exchange or other market that slows down trade executions, might give slower traders time to react to a change in the NBBO, and perhaps to change their quotes, which can serve to prevent their stale limit orders from being picked off by HFT traders.
9. In general, flash orders are communicated to certain market participants and either executed immediately or withdrawn immediately after communication. Immediate or cancel orders are generally displayed to the general public.
10. Flash orders can give HFTs the order flow information that they need to front run large incoming orders.
11. When the NBBO shifts, slow traders, whose quotes (limit orders) become stale (fail to reflect changes in supply and demand, becoming more beneficial relative to the NBBO), do not react, leaving them vulnerable to being picked off (executed rather than adjusted) by faster HFT traders.
12. Crumbling quotes begin as identical quotes on multiple exchanges at the NBBO, but one at a time, they rise or fall to a less aggressive range (falling bid or rising offer), until the NBBO itself becomes less aggressive (lower bid or higher offer).
13. Critics of the IEX, including Jeff Sprecher, CEO of the ICE, parent company of the NYSE, have claimed that the IEX primary peg order gives the exchange an unfair advantage because it does not display the trader's actual limit, while SEC rules require exchanges to display their best quotes.
14. A tick size of 0.000001 is practically meaningless. Any institution could develop an algorithm to create a meaningless improvement (called "stepping ahead" on any trader's quote) to free-ride off the quotes of others. Hence, any trader not employing her own algorithm would essentially have no practical way to post a quote that would not be instantly "jumped" by the tiniest increment. Rule 612 of Reg NMS prevents such "sub-penny" quotes (but not actual transactions prices) for shares selling for more than $1, which serves to protect meaningful limit orders, while still allowing for transactions within narrow spreads.
15. **a.** Loss of liquidity when the market needs it most leads to market fragility.
    b. The opposite of market fragility is market resilience or robustness, a scenario in which markets gain liquidity when they need it most.
    c. Most quotes can be withdrawn nearly instantly by HFT traders, which renders many quotes rather meaningless if they cannot be taken or executed by other investors. Thus, spreads might be wider than they appear. In fact, more than 95% of quotes in stock markets are ultimately canceled. For example, there is empirical evidence that some market liquidity and market spreads are illusory (Eggington, Van Ness, & Van Ness, 2016; Coffee, 2014), in large part because of flash trading

and quote stuffing, the potentially unethical practice of placing and withdrawing large numbers of quotes.

16. **a.** First, sponsored access is the practice of registered broker-dealers loaning their Market Participation Identification Numbers (MPIDs) to HFTs, enabling them to gain faster, direct access to markets by bypassing the broker's trading systems. "Naked access" occurs when HFT orders do not pass through the sponsor's pretrade risk controls.
    b. Algorithmic orders resulting from naked access can lead to runaway executions. Tracking down destabilizing naked access trades run amok can be tricky because exchanges cannot know where they originated, and the sponsoring brokerage firm must be involved in the identification.
    c. Short selling a security without actually borrowing it.
    d. Sellers might have difficulties acquiring shares to cover their short positions, creating counterparty risk, leaving their brokers and ultimately the relevant exchange and clearinghouse liable for covering short transactions.
17. An uninformed and unskilled individual investor would certainly encounter certain disadvantages trading in today's markets. This doesn't mean that she shouldn't invest, though. First, over the past century, buy-and-hold strategies have delivered decent long-term returns to stock market investors. There are ample, though sometimes expensive ways to obtain information-related assistance in the investment decision-making process. However, the trading outlook should be brighter for the better-trained and more informed experienced trader, even one who lacks superior technology. HFT often focuses on detecting information from the order flow and responding quickly to that information. If the experienced trader is careful not to reveal to HFT traders his useful information and analysis underlying his order flow, he should still be able to use his information and skills to compete effectively in stock markets despite lower speed and less sophisticated execution technology. Sophisticated technology is probably not a match for superior information as long as the superior information is not inadvertently shared in the trading process.

CHAPTER 5

1. **a.** While even insiders do not have perfect information concerning the stock valuation of their companies, their information is likely to be superior to that of the general public. If insiders are able to buy and sell shares on the basis of their inside information, they will convey this information to the general public through their trading activity and resulting price changes. The general public will buy and sell stock to reflect this information. Thus, share prices will more quickly and accurately reflect the best and most recent information relevant to the pricing of shares. Markets will more closely resemble the theoretical strong-form efficient market where prices fully reflect all information.
    b. If all traders were limited to trading on the basis of information only held by the general public, then no one would have an incentive to conduct costly research needed to uncover new information. There will be no trading based on new

information and market prices will not reflect new information. Society's resources will not be allocated on the basis of new information, reducing allocative efficiency.

2. Insider trading is not specifically mentioned in this rule. This rule is rather broad in its scope, and provides regulatory authorities and firm operating and compliance officers significant judgment in ensuring that anti-fraud objectives are being fulfilled.
3. **a.** While the law of insider trading derives from Section 10(b) of the Securities Exchange Act of 1934 and Rule 10b-5, neither the statute nor the rule addresses insider trading explicitly (except in the scenario of corporate acquisitions), leaving it to the court system to create legal doctrine about when insider trading should be pursued as a type of securities fraud.
   b. The SEC seeks to "level the playing field" and to maintain the integrity of markets. Thus, the SEC has historically viewed exploiting informational advantages via inside trading as being harmful to market integrity ("fairness" approach). However, the Supreme Court's rulings suggest that it views all information-based trading as reflecting informational advantages, in effect implying that all trading is based on informational advantage or perceived informational advantage. Instead, the Supreme Court seeks to limit illegality of insider trading to breaches of trust and confidence against firms ("property rights" approach). That is, if inside information is property that belongs to shareholders, insiders cannot misappropriate this information and legally trade on this inside information.
4. Dodd-Frank was intended to:
   1. promote the financial stability of the United States by improving accountability and transparency in the financial system;
   2. improve supervision of systemic risk, promote market discipline, and respond to threats to financial stability, in part by providing for the establishment of the Financial Stability Oversight Council (FSOC);
   3. end the informal "too big to fail" regulatory policies of the regulatory system;
   4. protect the American taxpayer by ending bailouts;
   5. protect consumers from abusive financial services practices, in part by improving transparency and creating the Consumer Financial Protection Board;
   6. end proprietary trading of bank assets by providing for the Volcker Rule, which also limits bank relationships with hedge funds and private equity firms; and
   7. force banks to maintain "skin in the game" by requiring sales of credit and mortgage-backed products to be accompanied by banks' retention of at least 5% of the instruments' credit risk unless the underlying loans meet certain standards that reduce risk.
5. Moral hazard is postcontractual opportunism in which the incentives or actions of one contracting party are not freely observable. The Volcker Rule limits banks and nonbank financial institutions supervised by the Fed with respect to proprietary trading and investment in and sponsorship of hedge funds and private equity funds. Trading exacerbates the moral hazard problem by allowing traders to impose potential losses on clients, insurers, creditors, and governments; the Volcker Rule prevents the government and its insurers from moral hazard problems caused by traders employed by banks.
6. Regulations can apply only to members who voluntarily join the regulated body. Regulations will not help entities (e.g., the general public) damaged by contracts in

markets if they are not participants of the regulated body, as harmed members will have no recourse against outside parties. Private regulators might be subject to more conflicts of interest than their government counterparts. Self-regulation can easily be created to serve as barriers to entry in the market, reducing competition. Private regulators do not have as much power to enforce regulations as does a government. For example, private regulators cannot imprison their members or enforce subpoena power.

7. The Securities Act of 1933 does not contain merit tests (e.g., tests as to whether the securities are valuable) as did earlier "blue skies" legislation, but instead provides for full disclosure of all material facts. This "sunlight theory of regulation" is based on the assumption that if investors are provided with all necessary and relevant information, they will make wise investment decisions. Presumably, firms that are forced to provide this information will pay attention to the merits of the securities that they sell.
8. **a.** Rules based: Do not exceed 55 miles per hour. Principles based: Do not drive faster than is necessary to maintain safety and fuel economy.
   b. In setting speed limits, rules-based regulations might need to specifically distinguish between driving conditions, road conditions, light conditions, vehicle type, driver experience, visibility, road surface conditions, traffic, and so on. Such conditions complicate the rule-setting process, and lead to more complicated sets of rules.
   c. Enforcement officers and traffic-court judges will need to interpret vague rules contingent on situations and conditions, and likely leave it to case law or drivers themselves to determine speed levels that are safe.
9. **a.** Consider a large block that executes under best price. It is likely that slices of the order would execute sequentially, at the best price then existing at each stage of the execution, without regard for likely slippage or reactions of competing traders' algorithms including quote cancellations. Essentially, the order steps through quotes, the national best bid or the national best offer sequentially until the order is fully executed. Under best execution, the broker would seek to get the best overall weighted average price, given appropriate considerations to the characters of the various markets, speed and certainty of execution, and making use of the broker's experience and algorithms with respect to getting the best overall price on the order. Problems with execution are exacerbated when the NBBO is for a small number of shares, say, 100, and reveals that more and larger orders are forthcoming from the same party. However, as a side issue, consider the Intermarket Sweep Order (ISO), an Immediate or Cancel limit order that allows the order to sweep the book (e.g., buy all the shares in the market with the original NBBO for a quantity of shares smaller than that of the order itself) without regard to any external quotes. This ISO allows for a work-around to the order protection rule (trade-through rule) of NMS, particularly when speed or certainty of execution is important.
   b. Best price order handling has the broker route the order or fraction thereof to the market with the best price, based on the prevailing NBBO, without regard for order size, sped or quote or market reliability. Best execution provides the broker more flexibility in order execution, allowing consideration for (1) the character of the market for the security; (2) the size and type of transaction; (3) the number of markets checked; (4) the accessibility of the quotation; and (5) the terms and

conditions of the order as communicated to the firm. The 5 considerations were taken from FINRA's best execution definition.

c. The not-held order type authorizes the broker to seek the best execution on behalf of her client, freeing the broker from Reg NMS best price obligations.

10. First, if the exchange mishandles either of its responsibilities poorly enough, it will fail, and it fails with respect in its duties to all of its stakeholders. As an SRO, the NYSE is overseen by the SEC; the SEC will simply impose regulations or restrictions on the NYSE if the NYSE fails in its SRO capacity. More generally, the NYSE can argue that its SRO responsibilities are essentially a product quality control mechanism. On the other hand, do consider that regulation is mandatory and applies to users of NYSE services. A government regulator has the power to impose civil and criminal penalties, essentially forcing all parties to adhere to its regulations. Should exchanges fail in their SRO capacities, their reputations will suffer, and new exchanges (e.g., the IEX) have the potential to emerge, possibly forcing less effective exchanges from the market. Also, bear in mind that many European countries, including the U.K. and France, make limited use of SROs and rely more on government regulation. As in the U.S. (e.g., NYSE and FINRA), China, Hong Kong, Singapore, and Sweden rely more on SROs as regulators, though not as much as Japan. Australia moved in 2010 from a strong reliance on SRO-based regulation (ASX) to a government-centered regulator (ASIC). So, there is quite a bit of variance in the profit-oriented/SRO model.

11. Listing on the NYSE will increase the profile of the listing company, among consumers and investors alike. Thus, there may be product market enhancements as well as investment benefits. Listing on the NYSE will enable the listing firm to be exposed to and receive coverage from U.S. financial analysts. In addition, listing on the NYSE and being subject to the U.S. regulatory environment signals to the investing public the firm's commitment to quality and regulatory compliance. In effect, the NYSE serves as a private market mechanism to certify investment, quality, and integrity standards.

12. Clawbacks are normally instituted after payment or vesting, whereas malus is normally instituted before payment or vesting.

13. Many traders do not post their quotes on the "lit" venues such as exchanges. Instead, in order to hide or camouflage their trading intentions, they frequently trade in dark liquidity pools, use non-displayed order types, or display only portions of actual order sizes. Exchanges and other market venues allow traders to use order types that obscure their best quotes, and many market venues allow traders to respond to flash orders, providing trading opportunities without posting any quotes at all. In addition, many quotes (limit orders) are easily canceled, so that their meaning that they are not reliable.

CHAPTER 6

1. a. The stock is valued as follows:
$P = \$2/(0.03 + 4 \times 0.01) = \$28.57$

b. $P = \$2/(0.03 + 4 \times 0.005) = \$40.00$

c. $P = \$2/(0.03 + 2 \times 0.005) = \$50.00$

d. $E[R] = 0.03 + 2 \times 0.005 = 0.04$

e. First, divide the proportional spread by 2 ($0.005/2 = 0.0025$) to obtain each segment, then adjust the price:
$P_b = \$50 \times (1 - 0.0025) = \$49.875$
$P_a = \$50 \times (1 + 0.0025) = \$50.125$

2. Perhaps there are many reasons for this, but the most relevant to this chapter is that members of the tribal society share in the catch of fish whereas modern fishermen compete against one another to harvest the catch. Similarly, traders compete against one another to acquire useful information, a scarce and valuable resource whose worth is diminished if used by another trader first.

3. Kyle's informed demand function x is as follows:

$$x = \frac{v - E[v]}{\sqrt{\frac{\Sigma_0}{\sigma_u^2}}} = \alpha + \beta v = -\frac{E[v]}{\sqrt{\frac{\Sigma_0}{\sigma_u^2}}} + \frac{v}{\sqrt{\frac{\Sigma_0}{\sigma_u^2}}}$$

where $\beta = 1/\sqrt{\frac{\Sigma_0}{\sigma_u^2}}$ and the expected value or price of the asset is $E[v] = p_0$. Thus, we can rewrite the informed trader's demand function as a function of the informed trader's per share profit as:

$$x = -\frac{E[v]}{\sqrt{\frac{\Sigma_0}{\sigma_u^2}}} + \frac{v}{\sqrt{\frac{\Sigma_0}{\sigma_u^2}}} = \beta(v - p_0).$$

4. a. Given:

$$v = 100; \quad \Sigma_0 = 60; \quad \sigma_u^2 = 10{,}000; \quad E[v] = 100$$

$$x = -\frac{E[v]}{\sqrt{\frac{\Sigma_0}{\sigma_u^2}}} + \frac{v}{\sqrt{\frac{\Sigma_0}{\sigma_u^2}}} = -\frac{100}{\sqrt{\frac{60}{10{,}000}}} + \frac{100}{\sqrt{\frac{60}{10{,}000}}} = 0$$

That is, if the private information does not change the informed trader's expected value of the stock from the market value, he will not trade. His demand for the stock at the current price is zero.

b. Given:

$$v = 100; \quad \Sigma_0 = 60; \quad \sigma_u^2 = 10,000; \quad E[v] = 90$$

$$x = \frac{v - E[v]}{\sqrt{\frac{\Sigma_0}{\sigma_u^2}}} - \frac{E[v]}{\sqrt{\frac{\Sigma_0}{\sigma_u^2}}} + \frac{v}{\sqrt{\frac{\Sigma_0}{\sigma_u^2}}} = -\frac{90}{\sqrt{\frac{60}{10,000}}} + \frac{100}{\sqrt{\frac{60}{10,000}}} = \frac{10}{0.07746} = 129.1$$

That is, the private information increases informed demand for the stock by or to 129.1 shares.

c. Given $\Sigma_0 = 60$, and $\sigma_u^2 = 10,000, E[\pi] = \sqrt{\Sigma_0 \sigma_u^2} = 774.60$, or 6 per share.

d. Assume that $E[v] = 90$, $\Sigma_0 = 60, \sigma_u^2 = 10,000, x = 129.1$ and $u = 0$.

First, we will calculate α and β:

$$\alpha = -\frac{E[v]}{\sqrt{\frac{\Sigma_0}{\sigma_u^2}}} = -\frac{90}{\sqrt{\frac{60}{10,000}}} = -1161.895$$

$$\beta = \frac{1}{\sqrt{\frac{\Sigma_0}{\sigma_u^2}}} = \frac{1}{\sqrt{\frac{60}{10,000}}} = 12.91$$

Next, we calculate the price set by the dealer:

$$\begin{aligned} p &= E[v] + \left[\frac{\beta\Sigma_0}{\beta^2\Sigma_0 + \sigma_u^2}\right][x + u - \alpha - \beta E[v]] \\ &= 90 + \left[\frac{12.91 \cdot 60}{12.91^2 \cdot 60 + 10,000}\right][129.1 + 0 + 1161.895 - 12.91 \cdot 90] \\ &= 90 + 0.0387 \cdot 129.1 = 95 \\ &\quad E[v] \qquad \lambda \quad x + u \end{aligned}$$

Similarly, since the dealer sets the price at $p = \mu + \lambda(x + u) = E[v] + \lambda(x + u)$, and λ is calculated as follows:

$$\lambda = \frac{1}{2}\sqrt{\frac{\Sigma_0}{\sigma_u^2}} = \frac{1}{2}\sqrt{\frac{60}{10,000}} = 0.0387$$

The dealer sets the stock price at $p = 90 + 0.0387(129.1 + 0) = 95$ based on the total demand for shares. Ultimately, the informed trader makes 5 per share buying from the dealer and the dealer makes 5 per share by buying shares from uninformed traders who pay 95 for shares that are known to informed traders to be worth 100. The dealer loses 5 per share relative to the 100 intrinsic value of the stock by selling to informed traders at 95.

5. a. The Kyle model suggests that informed traders have two ways to attempt to exploit their informational advantage. First, informed traders can choose to trade in large volume, earning small returns on large transactions as dealers adjust their spreads to reflect the suspected informed trader order volume. Second, informed traders could choose to trade in small volume, earning large returns on small transactions since dealers will not suspect informed trader presence. To avoid dealer spread adjustment, informed traders will tend to avoid large transactions. But many small transactions will also alert dealers. Thus, informed traders are rather likely to camouflage themselves by neither executing large transactions nor large numbers of transactions, and medium sized transactions are more likely to be undertaken by informed traders, and ultimately move markets.

b. Kyle (1985) predicts a linear relationship between order sizes (flow) and market impact.

c. Financial blogging reduces problems related to information asymmetries. This more equal access to information among investors will tend to reduce Kyle's lambda and reduce the market impact associated with order flow.

6. Unbalanced order arrivals indicate the presence or likelihood of informed trader orders. This causes the dealer to bias the bid downwards or offer upwards to reflect the possibility of increased informed trader activity.

7. a. The greater the probability that a trade has originated with an informed trader, the greater will be the spread. Increased information asymmetries lead to increased spreads. If all traders are informed and there are no noise traders, dealer spreads will be higher, because dealers cannot compete effectively against informed traders without noise traders. If all traders are noise traders, dealer spreads can be zero. When setting his quotes, the dealer needs to account for the probability that an informed trader will transact at his quote and will set his bid price P_b and ask price P_a as follows:

$$P_b = P_L + (1-\pi)\,(P_H + P_L)/2 \qquad P_a = \pi P_H + (1-\pi)(P_H + P_L)/2$$

Informed Traders:

$$P_b = P_L \qquad P_a = P_H$$

Uninformed Traders:

$$P_b = (P_H + P_L)/2 \qquad P_a = (P_H + P_L)/2$$

The spread is simply the difference between the ask and bid prices:

$$P_a - P_b = \pi\,(P_H - P_L)$$

Informed Traders:

$$P_a - P_b = P_H - P_L$$

Uninformed Traders:

$$P_a - P_b = 0$$

b. Price uncertainty is reflected in the Glosten and Milgrom model through the high and low prices, P_H and P_L. Greater differences between these two prices reflect greater price uncertainty.

8. Solve as follows: $P_a - P_b = \pi(P_H - P_L)$; $14.1 - 12.9 = \pi(3)$; $\pi = 0.4$.

9. a. Based on the standard Arrow-Pratt risk aversion model used in the Stoll inventory model, we first calculate the first derivative of utility with respect to wealth: $dU/dW = d\ln(W)/dW = 1/W$. Next, we calculate the second derivative of utility with respect to wealth: $d^2U/dW^2 = d(1/W)/dW = -1/W^2$. Now, we calculate ARA as follows: $\text{ARA} = -U''(W)/U'(W) = -(-1/W^2)/(1/W) = 1/W = 1/20{,}000$

b. $(P - P_b) = \frac{1}{2}(\sigma^2)ARA$; $(P - P_b) = \frac{1}{2}(10{,}000) \times 1/20{,}000 = 0.25 = (P - P_b)$; $P_b = 99.75$
$(P_a - P) = \frac{1}{2}(\sigma^2)ARA$; $(P_a - P) = \frac{1}{2}(10{,}000) \times 1/20{,}000 = 0.25 = (P_a - P)$; $P_a = 100.25$
$(P_a - P_b) = (\sigma^2)ARA = (10{,}000)/20{,}000 = 0.5 = 100.25 - 99.75$
The discount and premium are both $0.25. Notice that in the absence of inventory, the dealer spread is centered about the consensus price of 100.

c. $(P - P_b) = \sigma_{i,P}ARA + \frac{1}{2}(\sigma^2)ARA = 1000/20{,}000 + \frac{1}{2}(10{,}000)/20{,}000 = 0.3$; $P_b = 99.7$

d. $(P_a - P) = -\sigma_{i,P}ARA + \frac{1}{2}(\sigma^2)ARA = -1000/20{,}000 + \frac{1}{2}(10{,}000)/20{,}000 = 0.2$;
$P_a = 100.2$
$(P_a - P_b) = (\sigma^2)ARA = (10{,}000)/20{,}000 = 0.5 = 100.2 - 99.7$
The discount and premium are \$0.3 and \$0.2. Notice that in the presence of inventory, the dealer spread is no longer centered about the consensus price of 100, however the amount of the spread is unchanged.

10. Transactions costs tend to increase dealer spreads. In the absence of new information, execution prices often tend to bounce between bid and ask prices. For example, after a trade at the bid price, the next trade would be either at the same price or at the offer. After a trade at the offer, the next trade would be either at the same price or the bid. That is, price changes would tend to be either zero or at the opposite side of the spread. Thus, execution prices will tend to be either unchanged or to change in the opposite direction from the most recent transaction. Thus, transactions costs tend to induce negative serial correlation in asset prices.

11. In the event of a transaction at the bid, the risk-averse dealer would adjust quotes to promote inventory equilibrating trades. For example, after a transaction at the bid (which implies that the dealer purchased additional securities), both bid and ask quotes would fall so that the dealer would be able to liquidate this additional inventory, or at least not as readily increase it. That is, the dealer would want to discourage subsequent sales and to encourage purchases by traders so as to not significantly imbalance her portfolio. Similarly, after a transaction at the offer (which implies that the dealer sold additional securities), both bid and ask quotes would rise so that the dealer would be able to more easily purchase or replenish recently sold inventory, or at least not as readily further decrease it. That is, the dealer would want to discourage subsequent purchases and to encourage sales by traders. Over time, transactions prices and dealer quotes would tend to exhibit negative serial correlation.

12. a. If the trader's \$10.01 limit order executes, and if the stock price continues to decline, he will have to option to sell his shares to the institutional investor at \$10.00, losing \$0.01 per share. However, if the trader's \$10.01 limit order executes, and the stock price rebounds, the trader has unlimited profit potential.

b. The institutional investor's large limit order has provided a put option to the trader; the institutional investor's limit order to buy is essentially a short position on a put, provided for free to the trader. First, if the stock price drops to \$10.01 and eventually rebounds, the institutional investor will lose out on the trade (the trader took it at the higher limit) and not profit from the increase. Alternatively, the stock price can drop to \$10.00 and continue to drop, meaning that the institutional investor purchases the stock for \$10.00 (200,000 shares from the trader should he choose to sell) and loses money on the transaction, losing more as the stock price continues to drop.

13. Models discussed in this chapter shed light on this relationship between the dealer spread and security risk:

- In the Stoll (1978) model, increased dealer spreads will result from increased security uncertainty. Increased security uncertainty increases (decreases) the Arrow-Pratt risk premium associated with the security sales (purchase) price due to the dealer's risk aversion, as indicated by her ARA (absolute risk aversion coefficient). Increased uncertainty might also affect inventory size, which in itself will not affect the size of the spread (the difference between bids and offers), but will affect the levels of these quotes.

- In the Glosten and Milgrom (1985) model, the spread is a function of the likelihood that there exists an informed trader in the market and the uncertainty in the value of the traded asset. The greater the uncertainty in the value of the traded asset, and the greater the likelihood that a trade has originated with an informed trader, the greater will be the spread. Thus, increased spreads result from asymmetric information availability and the associated adverse selection problem.
- In the Copeland and Galai (1983) model, an options pricing model can be used to value the implied put and the implied call associated with the dealer spread. The values of both options increase as uncertainty increases. The implied premium associated with this dealer spread is paid by liquidity traders who trade without information. Thus, in the Copeland and Galai option-based model, the spread widens as the uncertainty of the security price increases.
- If the need for immediacy arises from uncertainty, the Demsetz (1969) argument might imply that spreads will increase as uncertainty or dealer risk increases.

14. a. Frictions such as transactions costs: Because with random arrival of buy and sell orders, price changes would tend to be either zero or at the opposite side of the spread, reflecting frictions rather than moving consistently either up or down. Prices exhibiting negative correlations reflect market frictions. For example, transactions costs cause consecutive stock price changes to fluctuate in opposite directions.

b. Random arrival of information: Because if price changes are solely a function of random news arrival, price changes will be randomly split between increases and decreases. Market inefficiencies, or market frictions such as transactions costs or information asymmetries, are the cause of positive or negative correlations in consecutive transactions prices.

c. Asymmetric information availability: Transactions executed at bid prices would cause permanent drops in prices to reflect negative information, and transactions executed at offer prices would cause permanent increases in prices to reflect positive information. If the distribution of information is asymmetric, prices will exhibit positive serial correlation as informed traders communicate their information through their trading activity.

CHAPTER 7

1. a. Yes: For every fixed number of cards t that have been dealt, the sum S_t of the values of the t dealt cards is a random variable. Furthermore, this random variable is indexed by time (technically, by the number of cards that have been dealt). The common probability space is the probability space that arises from the set of outcomes of all possible ways to deal 52 cards from the deck.

b. No: The probability of drawing a given point value really does depend on which cards were dealt before, not just on the total received before the third drawing. We can demonstrate this with just a single example, involving the probability of having 9 points after 3 draws: $P(S_3 = 9 \mid S_0 = 0, S_1 = 2, S_2 = 6) \neq P(S_3 = 9 \mid S_2 = 6)$ to justify that the Markov property is violated. This means that the probability of having 9 points after three draws where the first and second draw point totals are 2 and 6 is different from the probability of having 9 points after three draws where the second draw point total is 6. The first draw does affect the probability

of drawing a 3 in the third drawing. Note that $Z_t = S_t - S_{t-1}$ denotes the value of the tth card that is dealt. Thus, $P(S_3 = 9 \mid S_0 = 0, S_1 = 2, S_2 = 6) = P(Z_3 = 3 \mid Z_1 = 2, Z_2 = 4)$ and $P(S_3 = 9 \mid S_2 = 6) = P(Z_3 = 3 | Z_1 + Z_2 = 6)$.

We calculate that:

$$P(Z_3 = 3|Z_1 = 2,\ Z_2 = 4) = \frac{p(Z_1 = 2, Z_2 = 4, Z_3 = 4)}{p(Z_1 = 2, Z_2 = 4)} = \frac{\frac{4\times4\times4}{52\times51\times50}}{\frac{4\times4}{52\times51}} = \frac{2}{25}$$

In order for $Z_1 + Z_2 = 6$, either $Z_1 = 2$ and $Z_2 = 4$, or $Z_1 = 4$ and $Z_2 = 2$, or $Z_1 = 3$ and $Z_2 = 3$. Thus:

$$P(Z_3 = 3|Z_1 + Z_2 = 6) = \frac{p(Z_1 + Z_2 = 6, Z_3 = 3)}{p(Z_1 + Z_2 = 6)}$$
$$= \frac{P(Z_1 = 2, Z_2 = 4, Z_3 = 3) + P(Z_1 = 4, Z_2 = 2, Z_3 = 3) + P(Z_1 = 3, Z_2 = 3, Z_3 = 3)}{P(Z_1 = 2, Z_2 = 4) + P(Z_1 = 4, Z_2 = 2) + P(Z_1 = 3, Z_2 = 3)}$$

$$\frac{\frac{4\times4\times4}{52\times51\times50} + \frac{4\times4\times4}{52\times51\times50} + \frac{4\times3\times2}{52\times51\times50}}{\frac{4\times4}{52\times51} + \frac{4\times4}{52\times51} + \frac{4\times3}{52\times51}} = \frac{19}{275} \neq \frac{2}{25}$$

c. Yes: Because $E[S_t \mid S_{t-1}] = S_{t-1} + E[S_t - S_{t-1} \mid S_{t-1}]$ and $S_t - S_{t-1} > 0$, therefore, $E[S_t \mid S_{t-1}] > S_{t-1}$.

d. Yes: Because the cards are dealt with replacement, each card is dealt from the originally randomly shuffled deck. This implies that the choice of each card dealt is independent of the choice of any other card that is dealt. Thus, the random variables $\{Z_t\}$ are independent of one another. Thus, the resulting process is Markov.

e. Yes: By the same reason as in part c.

f. Yes: First, we write $E[S_t \mid S_{t-1}] = S_{t-1} + E[S_t - S_{t-1} \mid S_{t-1}]$. Note that because $Z_t = S_t - S_{t-1}$, $S_0 = 0$, and $S_{t-1} = Z_1 + \ldots + Z_{t-1}$, the random variable $S_t - S_{t-1}$ is independent of the random variable S_{t-1}. This implies that $E[S_t - S_{t-1} \mid S_{t-1}] = E[S_t - S_{t-1}] = E[Z_t]$. As the tth card is being dealt, the probability is 1/52 of its being any given card in the deck. There are $4 \times 5 = 20$ cards that have a point value of 1. There are $4 \times 3 = 12$ cards that have a point value of 0. There are $4 \times 5 = 20$ cards that have a point value of -1. Thus:

$$E[Z_t] = 1 \times \frac{20}{52} + 0 \times \frac{12}{25} + (-1) \times \frac{20}{52} = 0$$

This means that $E[S_t \mid S_{t-1}] = S_{t-1} + E[S_t\text{-}S_{t-1} \mid S_{t-1}] = S_{t-1}$. We conclude that S_t is a martingale.

2. a. First, note that observations from an arithmetic Brownian motion process are normally distributed, in this case with $\mu = .05$ and a standard deviation $\sigma = 5$. We calculate the probability of the range as follows:

$$\Pr\{.3 \le r_5 \le .5\} = N\{[\text{upper limit} - \mu T]/[T^{.5\sigma}]\} - N\{[\text{lower limit} - \mu T]/[T^{.5\sigma}]\}$$

$$= N\{[.5 - 5 \times .05]/[5^{.5 \times .5}]\} - N\{[.3 - (5 \times .05)]/[5^{.5 \times .5}]\} = .588468 - .517835 = .070633$$

b. $$\Pr\{ln(S_5 / S_0) > 2\} = 1 - N\{[ln(2) - T \times (\mu - .5\sigma^2)]/[T^{.5\sigma}]\}$$

$$= 1 - N([.63147 - 5 \times (.05 - .5 \times .25)]/[5^{.5 \times .5}]) = 1 - N(.95538) = .169693$$

c. Yes: The return process is a martingale because it reflects a constant drift, so its expected value $E[r_t] = \mu = .05$ for all t. However, because the drift is positive, the stock price process is a submartingale. Also, because this return process follows arithmetic Brownian motion, and arithmetic Brownian motion is always a martingale, the return process must be a martingale.

3. Since $S_t \sim N(50,4t)$, then $S_3 \sim N(50,12)$; that is, the distribution of the stock price in the third period has an expected value of 50 and a variance equal to $3 \times 4 = 12$. Using a standard z-table, we find that this probability equals $P(S_t < x) = P((S_3 - S_0)/\sigma)$:

$$P(S_3 < 56) = P\left(\frac{S_3 - 50}{\sqrt{12}} < \frac{56 - 50}{\sqrt{12}}\right) = P(Z < 1.73) = 0.9582.$$

4. a. $\overline{R}_R = 0.062$
$\overline{R}_B = 0.106$
$\overline{R}_M = 0.098$

b. $\sigma_R^2 = 0.000696$ (Population); 0.00087 (Sample—divide squared deviations by $n - 1$)
$\sigma_B^2 = 0.008824$ (Population); 0.01103 (Sample)
$\sigma_M^2 = 0.001576$ (Population); 0.00197 (Sample)

c. Assuming variance and correlation stability, the forecasted values would be the same as the historical values in parts a and b. Standard deviations are square roots of variances.

5. a. Since probabilities must sum to 1, the probability must equal 0.15.

b. First, note that there is a 0.25 probability that the return will be 0.05 (0.10 + 0.05 + 0.10) and 0.20 and 0.55 probabilities that the return will be 0.15. Thus, the expected return is $0.05 \times 0.25 + 0.10 \times 0.20 + 0.15 \times 0.55 = 0.115$. The variance is $0.25 \times (0.05 - 0.115)^2 + 0.20 \times (0.10 - 0.115)^2 + 0.55 \times (0.15 - 0.115)^2 = 0.001775$, which implies a standard deviation equal to 0.04213.

6. Standardize returns by standard deviations and consult "z" tables (such as the one in Mathematics Appendix A.4):
$(R_i - E[R])/\sigma = z$. Only use positive values for z here.

a. $(0.05 - 0.15)/0.10 = z(\text{low}) = 1$; $(0.25 - 0.15)/0.10 = z(\text{high}) = 1$
From the z table (Table A.1 on page 368), we see that the probability that the security's return will fall between 0.05 and 0.15 is 0.34. The value 0.34 is also the probability that the security's return will fall between 0.15 and 0.25. Therefore, the probability that the security's return will fall between 0.05 and 0.25 is 0.68.

b. From Table A.1 in the text appendix A.4, we see that the probability is 0.34.

c. 0.16
d. 0.0668

7. Simply reduce the standard deviations in the z scores in Problem 6 to 0.05.
a. 0.95
b. 0.47
c. 0.0228
d. 0.0013

8. a. The daily population variance is 0.003066. The daily sample variance is 0.00317.
b. The monthly population variance is $0.091982 = 0.003066 \times 30$. The monthly sample variance is $0.095 = 0.00317 \times 30$.
c. The Parkinson measure results in a variance estimate equal to 0.022465:

$$\sigma_p^2 = 0.361 \cdot \left[\ln\left(\frac{HI}{LO}\right)\right]^2 = 0.361 \cdot \left[\ln\left(\frac{38.5}{30}\right)\right]^2 = 0.022465$$

9. a. First, calculate returns as follows:

	Company X	**Company Y**	**Company Z**
Date	Return	Return	Return
1/09	–	–	–
1/10	0	0	0.00207
1/11	0.00249	0.00625	−0.00413
1/12	0	0.00621	−0.00207
1/13	0.00248	0.00617	−0.00207
1/14	−0.00248	0	0.00208
1/15	0.03980	0.04907	0.04158
1/16	0.00239	−0.00584	−0.02994
1/17	−0.00238	0.00588	0
1/18	0.00239	0.00584	0.00205
1/19	0.00238	−0.00581	0
1/20	−0.00238	0.00584	0

b. c. The table below calculates population standard deviations. Sample standard deviations would be 0.0120, 0.0148, and 0.0163.

Stock	**Average Return**	**Standard Deviation**
X	0.004064	0.011479
Y	0.006693	0.014150
Z	0.000869	0.015537

10. The variance estimate is computed as follows:

$$\sigma_p^2 = 0.361 \cdot \left[\ln\left(\frac{HI}{LO}\right)\right]^2 = 0.361 \cdot \left[\ln\left(\frac{50}{25}\right)\right]^2 = 0.173444$$

The standard deviation is the square root of this value, or 0.416466.

11. If investors have used the Black-Scholes options pricing model to evaluate this call, the following should hold:

$8.20 = 75 \times N(d_1) - 80 \times e^{-0.1 \times 0.5} \times N(d_2)$

$d_1 = \{\ln(75/80) + (0.1 + 0.5\sigma^2) \times 0.5\} \div \sigma\sqrt{0.5}$

$d_2 = d_1 - \sigma\sqrt{0.5}$

Thus, we wish to solve the above system of equations for σ. There exists no closed form solution for σ. Thus, we will substitute and iterate to search for a solution. We first arbitrarily select $\sigma_1 = 0.35$. We find that this estimate for sigma results in a value of 6.90 for c_0. Since this call price is less than the market value 8.20, we know that σ is larger than 0.35. Thus, we try a larger value for σ, repeating the process until finding that $\sigma = 0.411466$. We have estimated the implied value with a greater degree of accuracy than is needed for most applications.

12. Implied volatilities are given as follows:

a. $X = 40$; $\sigma = 0.2579$
b. $X = 45$; $\sigma = 0.3312$
c. $X = 50$; $\sigma = 0.2851$
d. $X = 55$; $\sigma = 0.2715$
e. $X = 60$; $\sigma = 0.2704$

These values are obtained through a process of substitution and iteration. That is, readers should select trial values for σ to substitute into the Black-Scholes formula, and then compute the trial call value. A closer trial value for the call to the actual market price leads to a closer computed volatility to its Black-Scholes implied value. Each reader will probably use a process that will differ at least slightly from those used by others.

13. These risks to arbitrage might be summarized as follows:

Implementation risk: arises because transactions might not be executed at all or might be executed at prices that differ from what the arbitrageur anticipated. Worse, one leg of the arbitrage transaction might fail to execute, leaving the arbitrageur unhedged in an undesirable portfolio.

Model risk: Sometimes arbitrageurs simply don't fully understand or appreciate the models that they trade on. This misunderstanding is called model risk and can arise from a failure to appreciate the differences between model assumptions and reality or simply misusing or failing to understand the model.

Basis risk: Arises from slow market adjustments to inconsistent pricing. Basis risk is exacerbated by margin calls, forced sell-outs, and forced buy-ins, potentially shortening the arbitrageur time horizon to recover from the unanticipated price change.

Counterparty risk: The potential that a trade counterparty fails to fulfill her side of a transaction. This risk is reduced by exchange trades. In some instances, potential

restrictions on trading (e.g., short-sale restrictions and ownership limits) or uncertainties with respect therewith, transaction costs, and taxes along with uncertainties associated therewith can make arbitrage risky.

CHAPTER 8

1. With matrices, the weights are found by solving for b as follows:

$$\begin{bmatrix} 50 & 80 & 110 \\ 50 & 80 & 1110 \\ 1050 & 1080 & 0 \end{bmatrix} \begin{bmatrix} \#_A \\ \#_B \\ \#_C \end{bmatrix} = \begin{bmatrix} 30 \\ 30 \\ 1030 \end{bmatrix}$$

$$\mathbf{CF} \times \mathbf{b} = \mathbf{p}_0$$

Alternatively, without matrices, we set up and solve the following system:
$50b_A + 80b_B + 110b_C = 30$
$50b_A + 80b_B + 1110b_C = 30$
$1050b_A + 1080b_B = 1030$
First, subtract the second equation from the first equation. We obtain the equation: $1000b_C = 0$. Next, solve for b_C: $b_C = 0$.
Now, substitute $b_C = 0$ in the original second equation:
$50b_A + 80b_B = 30$
Now, we have an equation with two unknowns. Combine with our original third equation:
$50b_A + 80b_B = 30$
$1050b_A + 1080b_B = 1030$
Now, solve this 2×2 system for b_A and b_B. First, solve the first equation for b_A:
$50b_A = 30 - 80b_B$
$b_A = 0.6 - 1.6b_B$. Divide both sides by 50.
Now, substitute this equation for b_A in the second equation:
$1050(0.6 - 1.6b_B) + 1080b_B = 1030$
Simplify by combining similar terms:
$600b_B = -400$
Now, we know b_B:
$b_B = -0.666666$
We substitute b_B into our original third equation to obtain b_A:
$1050b_A + 1080 \times -0.666666 = 1030$
Finally, we obtain b_A:
$b_A = 1.666666$
We find that $b_A = 1.666666$, $b_B = -0.666666$, and that $b_C = 0$. This means that bond D is replicated by a portfolio consisting of 1.666666 of bond A and -0.666666 of bond B.

2. a. Its annual interest payments:
$i_y = \text{Int}/F$
$\text{Int} = i_y(F)$; $= (0.12)(1000) = \$120$

b. Its current yield: $cy = \text{Int}/P_0 = 120/1200 = 0.10$

c. Through substitution, we find yield to maturity to be 0.04697429 or 4.697429%.

3. a. Its annual interest payments: \$120, or \$60 every six months.

b. $120 \div 1200 =$ current yield $= 0.10$ or 10%.

c. Its yield to maturity y is found by substitution and eventually arriving at:

$$-1200 + 60/[1 + (y/2)]^1 + 60/[1 + (y/2)]^2 + \cdots + 60/[1 + (y/2)]^5 + 1060/[1 + (y/2)]^6$$

$y = 0.0476634$

4. The following matrix system may be solved for b to determine exactly how many of each of the bonds are required to satisfy the fund's cash flow requirements:

$$\begin{bmatrix} 1100 & 100 & 110 & 120 \\ 0 & 1100 & 110 & 120 \\ 0 & 0 & 1110 & 120 \\ 0 & 0 & 0 & 1120 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} = \begin{bmatrix} 30,000,000 \\ 15,000,000 \\ 25,000,000 \\ 35,000,000 \end{bmatrix}$$

$$\mathbf{CF} \times \mathbf{b} = \mathbf{P}_0$$

First, we invert matrix **CF** to obtain $\mathbf{CF}^{-1}$:

$$\begin{bmatrix} 0.000909 & -0.000083 & -0.00008 & -0.000079 \\ 0 & 0.000909 & -0.00009 & -0.000087 \\ 0 & 0 & 0.00090 & -0.000096 \\ 0 & 0 & 0 & 0.000892 \end{bmatrix}$$

$$\mathbf{CF}^{-1}$$

We then post-multiply $\mathbf{CF}^{-1}$ by $\mathbf{P}_0$ to obtain **b**. Alternatively, without matrices, we set up and solve the following system:

$1100b_1 + 100b_2 + 110b_3 + 120b_4 = 30{,}000{,}000$

$1100b_2 + 110b_3 + 120b_4 = 15{,}000{,}000$

$1110b_3 + 120b_4 = 25{,}000{,}000$

$1120b_4 = 35{,}000{,}000$

First, solve the fourth equation, we obtain b_4:

$b_4 = 31{,}250$

Now, substitute $b_4 = 31{,}250$ in the original third equation:

$1110b_3 + 120 \times 31250 = 25{,}000{,}000$

Solving this equation, we find that $b_3 = 19{,}144.14$. Now, substitute b_3, b_4 into the original second equation:

$1100b_2 + 110 \times 19{,}144.14 + 120 \times 31{,}250 = 15{,}000{,}000$

Solve this equation. We know that $b_2 = 8312.858$. Substitute b_2, b_3, b_4 into the original first equation:

$1100b_1 + 100 \times 8312.858 + 110 \times 19{,}144.14 + 120 \times 31{,}250 = 30{,}000{,}000$

Finally, we find that $b_1 = 21{,}193.5$. Note that we were able to use the bootstrapping method to solve his system. We find that the purchase of 21,193.5 bonds 1, 8312.858 bonds 2, 19,144.14 bonds 3, and 31,250 bonds 4 satisfy the insurance company's exact matching requirements.

5. a.i. First, find the yield to maturity (ytm) of the bond:

$$0 = NPV = \sum_{t=1}^{n} \frac{CF_t}{(1+ytm)^t} - P_0; \text{ yield to maturity} = ytm$$

$$0 = NPV = \sum_{t=1}^{n} \frac{1000}{(1+ytm)^1} - 900; \text{ solve for } ytm$$

$ytm = 0.111$

a.ii. Use ytm from part i in the duration formula:

$$Dur = \frac{\sum_{t=1}^{n} t \cdot \frac{CF_t}{(1+ytm)^t}}{P_0}$$

Note: Negative signs are omitted.

$$Dur = \frac{1 \cdot \frac{1000}{1.111}}{900} = Dur = 1 \text{ year}$$

b.i.

$$0 = NPV = \sum_{t=1}^{n} \frac{CF_t}{(1+ytm)^t} - P_0 = \frac{1000}{(1=ytm)^t} - 800$$

$ytm = 0.118$

b.ii.

$$Dur = \frac{\sum_{t=1}^{n} t \cdot \frac{CF_t}{(1+ytm)^t}}{P_0} = \frac{2 \cdot \frac{1000}{1.118^2}}{800} = 2$$

c. $ytm = 0.126$;

$$Dur = \frac{3 \cdot \frac{2000}{(1.126)^3}}{1400} = 3$$

d. There are several ways to work this problem. First, consider the cash flows of the portfolio:

$$P_0 = 900 + 800 + 1400 = 3100$$

$$CF_1 = 1000; CF_2 = 1000; CF_3 = 2000$$

$$0 = NPV = \frac{1000}{(1+ytm)^1} + \frac{1000}{(1+ytm)^2} + \frac{2000}{(1+ytm)^3} - 3100; ytm = 0.122$$

$$Dur = \frac{\sum_{t=1}^{n} t \cdot \frac{CF_t}{(1+ytm)^t}}{P_0} = \frac{\frac{1 \cdot 1000}{1.122} + \frac{2 \cdot 1000}{(1.122)^2} + \frac{3 \cdot 2000}{(1.122)^3}}{3100}$$

$Dur = 2.161$ years

Second, note that the portfolio duration is a weighted average of the bond durations:
$(900/3100) \times 1 + (800/3100) \times 2 + (1400/3100) \times 3 = 2.161$

6. The duration of a pure discount bond equals its maturity.

7. a.

First, find the bond's *ytm*:

$$0 = NPV = \frac{100}{(1+ytm)^1} + \frac{100}{(1+ytm)^2} + \frac{100+1000}{(1+ytm)^3} - 900;\ ytm = 0.143$$

Now, use *ytm* to find duration:

$$Dur = \frac{\frac{1 \cdot 100}{1.143} + \frac{2 \cdot 100}{(1.143)^2} + \frac{3 \cdot 1100}{(1.143)^3}}{900} = 2.722$$

b.

$$0 = NPV = \frac{120}{(1+ytm)^1} + \frac{120}{(1+ytm)^2} + \frac{1120}{(1+ytm)^3} - 900;\ ytm = 0.165$$

$$Dur = \frac{\frac{1 \cdot 120}{1.165} + \frac{2 \cdot 120}{(1.165)^2} + \frac{3 \cdot 1120}{(1.165)^3}}{900} = 2.672$$

c.

$$0 = NPV = \frac{100}{(1+ytm)^1} + \frac{100}{(1+ytm)^2} + \frac{100}{(1+ytm)^3} + \frac{1100}{(1+ytm)^4} - 900$$

$ytm = 0.134$

$$Dur = \frac{\frac{1 \cdot 100}{1.165} + \frac{2 \cdot 100}{(1.165)^2} + \frac{3 \cdot 100}{(1.165)^3} + \frac{4 \cdot 1100}{(1.134)^4}}{900} = 3.456$$

d.

$$0 = NPV = \frac{100}{(1+ytm)^1} + \frac{100}{(1+ytm)^2} + \frac{1100}{(1+ytm)^3} - 800$$

$ytm = 0.194$

$$Dur = \frac{\frac{1 \cdot 100}{1.194} + \frac{2 \cdot 100}{(1.194)^2} + \frac{3 \cdot 1100}{(1.194)^3}}{800} = 2.703$$

8. a. $\%\Delta P_0 = Dur \times \%\Delta(1+r);\ \%\Delta P_0 = 2.722 \times 0.10 = 0.2722$
$\%\Delta P_0 = 0.2722;\ \Delta P_0 = 0.2722 \times 900 = 244.98$

The new price is 900 + 244.98 = 1144.98

b. $\%\Delta P_0 = 2.672 \cdot 0.10 = 0.2622$; $\Delta P_0 = 235.98$; price = 1135.98

c. $\%\Delta P_0 = 3.456 \cdot 0.10 = 0.3456$; $\Delta P_0 = 311.04$; price = 1211.04

d. $\%\Delta P_0 = 0.2703$; $\Delta P_0 = 216.24$; new price = 1016.24

9.

$$Dur = 20772 \cdot \frac{900}{3500} + 2.672 \cdot \frac{900}{3500} + 3.456 \cdot \frac{900}{3500} + 2.703 \cdot \frac{800}{3500} = 2.893$$

10. Durations and convexities are as follows:

a. *Dur* = 4.203743015; Con = 20.31015, calculated as follows:

Dur A	Con A
92.59259258	158.7664481
171.467764	441.0179114
238.1496722	816.6998358
294.0119409	1260.339253
3743.207581	19255.18302

Add and then divide by $P_0 = 1079.8542$ to obtain:

4.203743015	20.31015527

b. *Dur* = 4.0373493; Con = 17.86, calculated as follows:

Dur B	Con B
107.1428571	170.8272595
191.3265306	457.5730164
256.2408892	817.0946722
305.0486776	1215.914691
3177.590392	15198.93363

Add and then divide by $P_0 = 1000$ to obtain:

−4.037349347	17.86034327

11. **a.** $P_{1A} = 995.7906904 = 1079.8542 - 4.203743 \times [(0.1 - 0.08)/(1.08)] \times 1079.8542$

$P_{1B} = 1072.095524 = 1000 - 4.037349 \times [(0.1 - 0.12)/(1.12)] \times 1000$

b. $P_{1A} = 1000.1770910 = 1079.8542 - 4.203743 \cdot [(0.1 - 0.08)/(1.08)] \times 1079.8542 + 0.5 \times 1079.8542 \times 20.3101 \times (0.1 - 0.08)^2$
$P_{1B} = 1075.667592 = 1000 - 4.037349 \times (0.1 - 0.12)/(1.12) \times 1000 + 0.5 \times 1000 \times 17.86 \times (0.1 - 0.12)^2$

c. $P_{1A} = 1000.00 = (100/0.1) \times (1 - 1/(1.1)^5) + 1000/(1.1)^5$
$P_{1B} = 1075.815735 = (120/0.1) \times (1 - 1/(1.1)^5) + 1000/(1.1)^5$
The new bond values given in part c are more precise. Note how much better the bond convexity model in part b estimates revised bond prices than the duration model in part a.

12. According to the pure expectations theory, we compute the two-year spot rate as follows:

$$(1+y_{0,2})^2 = \prod_{t=1}^{2}(1 + y_{t-1,t}) = (1 + 0.05)(1 + 0.08) = 1.134$$
$$y_{0,2} = [(1+0.05)(1+0.08)]^{1/2} - 1 = \sqrt{1.134} - 1 = 0.0648944$$

13. The three-year rate is based on a geometric mean of the short-term spot rates as follows:

$$(1+y_{0,3})^3 = \prod_{t=1}^{3}(1 + y_{t-1,t}) = (1 + 0.05)(1 + 0.06)(1 + 0.07) = 1.19091$$
$$y_{0,3} = [(1+0.05)(1+0.06)(1+0.07)]^{1/3} - 1 = \sqrt[3]{1.19091} - 1 = 0.0599686$$

14. The three-year rate is based on a geometric mean of the short-term spot rates as follows:

$$(1+y_{0,3})^3 = (1.07)^3 = 1.22504 = \prod_{t=1}^{3}(1 + y_{t-1,t}) = (1 + 0.05)(1 + 0.07)(1 + y_{2,3})$$

We solve for $y_{2,3}$ as follows:

$$1.22504 \div [(1 + 0.05)(1 + 0.07)] - 1 = y_{2,3} = 0.0903$$

15. **a.** First, we set up discount functions used by investors to value the bonds:
$1010D_1 = 1005$
$50D_1 + 1050D_2 = 1040$
$40D_1 + 40D_2 + 1040D_3 = 1020$
$40D_1 + 40D_2 + 40D_3 + 1040D_4 = 990$
Next, solve for, in order, D_1, D_2, D_3, and D_4:
$D_1 = 1005/1010 = 0.99505$
$D_2 = (1040 - 50 \times 0.99505)/1040 = 0.943093$
$D_3 = (1020 - 40 \times 0.99505 - 40 \times 0.943093)/1020 = 0.906225$
$D_4 = (990 - 40 \times 0.99505 - 40 \times 0.943093 - 40 \times 0.906225)/990 = 0.842524$
Finally, solve for $y_{0,1}$, $y_{0,2}$, $y_{0,3}$ and $y_{0,4}$:
$y_{0,1} = 1/D_1 - 1 = 1/0.99505 - 1 = 0.004975$
$y_{0,2} = (1/D_2)^{1/2} - 1 = (1/0.943093)^{1/2} - 1 = 0.029729$
$y_{0,3} = (1/D_3)^{1/3} - 1 = (1/0.906225)^{1/3} - 1 = 0.033367$
$y_{0,4} = (1/D_4)^{1/4} - 1 = (1/0.842524)^{1/4} - 1 = 0.043769$

b. Use $y_{0,}n$ and $y_{0,1}$ to solve for each $y_{1,n}$ as follows:
$y_{1,2} = (1 + y_{0,2})^2/(1 + y_{0,1}) - 1 = 0.055092$
$y_{1,3} = [(1 + y_{0,3})^3/(1 + y_{0,1})]^{1/2} - 1 = 0.047852$
$y_{1,4} = [(1 + y_{0,4})^4/(1 + y_{0,1})]^{1/3} - 1 = 0.05703$

c. Use $y_{0,n}$ and $y_{0,2}$ to solve for each $y_{2,n}$ as follows:
$y_{2,3} = [(1 + y_{0,3})^3/(1 + y_{0,2})^2] - 1 = 0.040683$
$y_{2,4} = [(1 + y_{0,4})^4/(1 + y_{0,2})^2]^{1/2} - 1 = 0.058001$

d. Use $y_{0,4}$ and $y_{0,3}$ to solve for $y_{3,4}$ as follows:
$y_{3,4} = [(1 + y_{0,4})^4/(1 + y_{0,3})^3] - 1 = 0.075607$

16. a. First, assume that bonds are priced with the following present value functions:

Equation Set A

$$964.3227 = 70D_1 + 70D_2 + 1070D_3$$
$$1010.031 = 80D_1 + 1080D_2 + 0D_3$$
$$938.4063 = 60D_1 + 60D_2 + 1060D_3$$

We will call this first set of equations Equation set A. This set contains three equations with three unknown values for which we need solutions. Note that the coefficient for D_3 in the second equation in Set A equals zero. If we can produce another equation with a zero coefficient for D_3, we will be able to solve for two unknowns in a two-equation system. We will call this next set of equations Equation Set B. If we multiply Equation 1A by (1060/1070), the coefficient for D_3 Equation 1B will match the coefficient for D_3 in Equation 3A (which will be rewritten as Equation 2B).

Equation Set B

$69.34579D_1 + 69.34579D_2 + 1060D_3 = 955.31031060/1070 \times Equation\ 1A = Equation\ 1B$
$60D_1 + 60D_2 + 1060D_3 = 938.4063 \qquad Equation\ 3A = Equation\ 2B$

Next, subtract Equation 2B from Equation 1B to eliminate D_3 and its coefficient 1060. The result of this subtraction is the first equation in Equation Set C. The second equation in Equation Set C is the second equation in Equation Set A multiplied by (9.345794/80). This multiplication will enable us to subtract later to eliminate unknown D_1:

Equation Set C

$9.345794D_1 + 9.345794D_2 + 0D_3 = 16.90401 \quad Equation\ 1B - Equation\ 2B = Equation\ 1C$
$9.345794D_1 + 126.1682D_2 + 0D_3 = 117.9943 \quad Equation\ 2A \times (9.345794/80) = Equation\ 2C$

When we subtract the second equation in Equation Set C from the first, Variable D_1 will be eliminated:
$0D_1 + -116.822D_2 + 0D_3 = -101.09 \qquad Equation\ 1C - Equation\ 2C$
Now, we can easily solve for $D_2 = -101.09/(-116.822) = 0.865333$. Now, using Equation 1C, we can solve for $D_1 = (16.90401 - 9.345794 \times D_2)/9.345794 = 0.943396$. Finally, we use Equation 1A to solve for $D_3 = (964.3227 - 70 \times D_1 - 70 \times D_2)/1070 = 0.782908$. Now that we have solved for our three discount functions D_1, D_2, and D_3, we can obtain spot rates:

$$\frac{1}{D_1} - 1 = 0.060 = y_{0,1}$$

$$\frac{1}{D_2^{1/2}} - 1 = 0.075 = y_{0,2}$$

$$\frac{1}{D_3^{1/3}} - 1 = 0.085 = y_{0,3}$$

Obtain the one- and two-year forward rates on loans originated in year 1 as follows:

$$y_{1,2} = \frac{(1+0.075)^2}{(1+0.06)} - 1 = 1.0902123 - 1 = 0.0902123$$

$$y_{1,3} = \sqrt{\frac{(1+0.085)^3}{1+0.06}} - 1 = \sqrt{1.2049897} - 1 = 0.097720228$$

The one-year forward rate on a loan originated in two years is computed as follows:

$$y_{2,3} = \frac{(1+0.085)^3}{(1.075)^2} - 1 = 1.052799 - 1 = 0.1052799$$

17. a. First, solve the following system for the discount functions d:

$$\begin{bmatrix} 50 & 50 & 1050 \\ 80 & 80 & 1080 \\ 110 & 1110 & 0 \end{bmatrix} \begin{bmatrix} D_1 \\ D_2 \\ D_3 \end{bmatrix} = \begin{bmatrix} 878.9172 \\ 955.4787 \\ 1055.4190 \end{bmatrix}$$

$$\mathbf{CF} \times \mathbf{d} = \mathbf{p_0}$$

We find that $D_1 = 0.943396$, $D_2 = 0.857338$, and $D_3 = 0.751314$. The spot rates are obtained as follows:

$$\frac{1}{D_1} - 1 = \frac{1}{0.943396} - 1 = 0.06$$

$$\frac{1}{D_2^{\frac{1}{2}}} - 1 = \frac{1}{0.857338^{\frac{1}{2}}} - 1 = 0.08$$

$$\frac{1}{D_3^{\frac{1}{3}}} - 1 = \frac{1}{0.751314^{\frac{1}{3}}} - 1 = 0.10$$

b. The weights are found by solving for **w** as follows:

$$\begin{bmatrix} 50 & 80 & 110 \\ 50 & 80 & 1110 \\ 1050 & 1080 & 0 \end{bmatrix} \begin{bmatrix} w_A \\ w_B \\ w_C \end{bmatrix} = \begin{bmatrix} 30 \\ 30 \\ 1030 \end{bmatrix}$$

$$\mathbf{CF} \cdot \mathbf{w} = \mathbf{p_0}$$

$$\underbrace{\begin{bmatrix} -0.03996 & 0.00396 & 0.002666 \\ 0.03885 & -0.00385 & -0.001660 \\ -0.00100 & 0.00100 & 0 \end{bmatrix}}_{\mathbf{CF^{-1}}} \cdot \underbrace{\begin{bmatrix} 30 \\ 30 \\ 1030 \end{bmatrix}}_{\mathbf{p_0}} = \underbrace{\begin{bmatrix} w_A \\ w_B \\ w_C \end{bmatrix}}_{\mathbf{w}}$$

We find that $w_A = 1.666666$, $w_B = -0.666666$, and $w_C = 0$, which means that bond D is replicated by a portfolio consisting of 1.666666 of Bond A and -0.666666 of Bond B.

18. USD2.30 × (CAD1/USD0.64) = USD2.30 × 1.5625 = CAD3.59375

19. a. CHF1.5 ÷ (CAD1/USD0.64) = CHF1.5/1.5625 = CHF0.96/CAD1
CHF9.6 are required to purchase 10 Canadian dollars. In Switzerland, one Big Mac should cost CHF1.5 per USD times 2.30 USD per Big Mac or CHF3.45 when purchase power parity holds.

20. a. The value of $1 is equal to the value of £0.6. Thus, the value of £1 must be $1/0.6 = $1.6667. Since the value of $1 is ¥108, the value of £1 must be 1.6667 × ¥108 = ¥180. Thus, we could have solved as follows: £1 = ($1 ÷ 0.6) × ¥108 = ¥180.
The value of $1 is equal to the value of ¥108. Thus, the value of ¥1 must be $1/¥108 = $0.0092593. Since the value of £1 is $1.6667, the value of ¥1 must be $1 ÷ (1.6667 × ¥108) = £0.0055556. Thus, we could have solved as follows:
¥1 = $1 ÷ (1/0.6 × ¥108) = £0.0055556.
$300 = £0.6 × 300 = £180; $300 = ¥108 × 300 = ¥32,400

21. First, we will consider the alternative of doing nothing.

Unhedged alternative:
Strategy: Wait six months and then sell ¥15,000,000 for dollars at the then-prevailing spot rate.
Result: All ¥15,000,000 is at risk. The expected value of the transaction is $144,230.77.

Forward market hedge:
Strategy: Sell ¥15,000,000 forward for dollars at once.
Result: $144,230.77 will certainly be received in six months. Transactions costs at time zero will total $500. Forgone interest over six months totals $12.50. The total amount (net of transactions costs and forgone interest) to be received in six months is $143,718.27. This amount is certain.

Money market hedge:
Strategy: Borrow ¥14,354,067 in Japan for six months at 9% per year; exchange ¥14,354,067 for $136,705.40 now; invest $136,705.40 for six months at 5% per year.
Result: Yen loan is repaid by receipts from sale in six months. $140,123.03 are obtained from U.S. investment. This amount is certain in the absence of default risk. Note that this strategy is inferior to the forward market hedge.

CHAPTER 9

1. a. Shareholders have at their disposal a variety of risk and portfolio management techniques to manage risk and don't need corporate managers to do so for them. For example, shareholders can hold large, well-diversified portfolios or take positions in a wide variety of highly liquid derivative securities to manage a variety of specific types of risk. Furthermore, when shareholders are left to themselves to manage the risks of their own portfolios, they can more easily fit their portfolios to their personal risk preferences than managers can.

b. When corporations take risks, they subject themselves to bankruptcy risks and the wide variety of associated direct bankruptcy costs (e.g., legal, accounting, and court fees, as well as reorganization and liquidation costs) and indirect costs of bankruptcy (e.g., inability to transact with and loss of goodwill from clients and suppliers). Shareholders are normally not in a position to evaluate pertinent information as are managers, or to be able to easily and cost-effectively access financial institutions and contracts needed to mitigate risks.

2. a. $c_T = \text{MAX}[0, S_T - X]$; $c_T = \$0$ or $\$15$

b. $\$100/\$90 - 1 = 0.1111$

c.

$$\alpha = \frac{c_u - c_d}{S_0(d - u)}$$

$$\alpha = \frac{15 - 0}{50(0.6 - 1.4)} = -0.375$$

d.

$$c_0 = \frac{-(1 + r_f)\alpha S_0 + C_d + \alpha d S_0}{(1 + r_f)}$$

$$c_0 = \frac{(1 + 0.1111) \cdot 0.375 \cdot 50 + 0 - 0.375 \cdot 0.6 \cdot 50}{(1 + 0.1111)} = 8.625$$

e. $p_0 = c_{0+} X/(1 + r_f)^T - S_0 = 8.625 + 55/(1.111111) - 50 = 8.125$

3. First, find the hedge ratio:

$$\alpha = \frac{c_u - c_d}{S_0(d - u)}$$

$$\alpha = \frac{8 - 2}{12(.83333 - 1.3333)} = -1$$

Now, value the call:

$$c_0 = \frac{-(1+r_f)\alpha S_0 + c_d + \alpha d S_0}{(1+r_f)}$$

$$c_0 = \frac{-(1+0.125)\cdot(-1)\cdot 12 + 2 - 1\cdot 0.83333\cdot 12}{(1+0.125)} = 4.88889$$

4. The statement for this problem leaves open a few ambiguities, primarily how jumps and riskless returns should be compounded over the course of the relevant time periods. Thus, depending on compounding methodologies, there should be several ways to complete each of these problems, but all should yield fairly similar solutions. First, we will estimate from the annualized volatility (σ) values for u, d, and p in the two-, three-, and eight-time-period cases:

$$u = e^{0.6\times\sqrt{0.75/2}} = 1.444; \quad u = e^{0.6\times\sqrt{0.75/3}} = 1.350; \quad u = e^{0.6\times\sqrt{0.75/8}} = 1.202$$

$$d = \frac{1}{1.444} = 0.692; \quad d = \frac{1}{1.35} = 0.741; \quad d = \frac{1}{1.202} = 0.832$$

$$p = \frac{e^{0.08\cdot 0.75} - 0.692}{1.444 - 0.692} = 0.449689; \quad p = \frac{e^{0.08\cdot 0.75} - 0.741}{1.35 - 0.741} = 0.564; \quad p = \frac{e^{0.08\cdot 0.75} - 0.832}{1.202 - 0.832} = 0.475$$

For the two-time-period framework, call valuation calculations proceed as follows:

$$a = INT\left[\text{MAX}\left[\frac{\ln\left(\frac{80}{50\cdot 0.692^2}\right)}{\ln\left(\frac{1.444}{0.692}\right)}, 0\right]\right] + 1 = 2$$

$$c_0 = \frac{0.449^2 \times 0.551^{2-2} \times [1.444^2 \times 0.692^{2-2} \times 50 - 80]}{(1+0.08\cdot 0.75/2)^2} = 4.62$$

Put values are found with put-call parity. The following are call and put values for the two-, three-, and eight-period frameworks:

n	c_0	p_0
2	4.62	30.13
3	3.91	29.42
8	3.86	29.37

The calls are easily valued in the two- and three-step models because the maximum number of up-jumps (two and three) are required for exercise. Call value calculations are summarized as follows for the three-period framework, where calls will not be exercised with two or fewer up-jumps:
$[0.0965 \times (122.98015 - 80)]/(1 + 0.08 \times 0.75/3)^3 = 3.91$
In the eight-period framework, at least 6 up-jumps were required for exercise. Call value calculations are summarized as follows for the eight-period framework, where calls will not be exercised with 5 or fewer up-jumps:

$$\{[0.0883 \times (104.258 - 80)] + [0.02279 \times (150.55 - 80)] + [0.00257 \times (217.395 - 80)]\}/(1 + 0.08 \times 0.75/8)^8 = 3.86$$

5. $S_0 = 50$, $X = 60$, $T = 0.75$, $r_f = 0.081$, $u = 1.2776$, $d = 0.7828$

$$p = \frac{(1 + r_f) - d}{u - d} = \frac{(1 + .081) - .7828}{1.2776 - .7828}$$

$p = 0.501295527$, $(1 - p) = 0.498704473$, $a = 2$, $T/n = 0.375$, $r_f \times T/n = 0.030375$ (adjust r_f)

$$c_0 = \frac{c_{u^2}p^2 + 2p(1-p)c_{ud} + (1-p)^2 c_{d^2}}{(1+r_f)^2} = \frac{21.61309 \times .251297 + 0 + 0}{(1+.030375)^2}$$

$c_0 = \$5.10$; based on put-call parity, $p_0 = \$11.61$

6. a. We are working with a binomial environment with two time periods in a single year with $u = 1.2$ and $d = .8333$. We can easily find the probability of upward movement to be $q = ([1 + \mathrm{r}] - d)/(u - d) = .1667/.3667 = .4545$. Now, we can use the binomial model to value the call:

$$c_0 = \frac{\sum_{j=0}^{2} \frac{2!}{j!(2-j)!} \times .4545^j (1-.4545)^{2-j} \mathrm{MAX}\left[1.2^j \times .8333^{2-j} \times 50 - 60, 0\right]}{(1+.0)^2}$$

$$= \frac{c_{2;u\ u}\, q^2}{(1+0)^2} = \frac{12 \times .2066}{1} = 2.48$$

Using the put-call parity relation to value the put, we find that the put is worth $p_0 =$ **12.48**.

b. Working with the same model as above, we want the probability of at least one up-jump in the stock's price over two periods, which equals one minus the probability of two down-jumps $= 1 - (1 - q)^2 = 1 - .545^2 =$ **.7025**.

7. a. $d_1 = 0.6172$; $d_2 = 0.1178$; $N(d_1) = 0.7314$; $N(d_2) = 0.5469$

$c_0 = 11.05$; with put−call parity: $p_0 = 4.34$

b. Use $X = 30$; $d_1 = 0.925$; $d_2 = 0.4245$; $N(d_2) = 0.6644$

$1 - N(d_2) = 0.3356$

8. Value the calls using the Black-Scholes model:

$c_0 = S_0 N(d_1) - X\mathrm{e}^{-rT} N(d_2)$

$d_1 = [\ln(S \div X) + (r + 0.5\sigma^2)T] \div \sigma\sqrt{T}$

$d_2 = d_1 - \sigma\sqrt{T}$

Thus, we will first compute d_1, d_2, $N(d_1)$, and $N(d_2)$ for each of the calls; and then we will compute each call's value. We will then use put−call parity to value each put. First find for each of the 15 calls values for d_1:

X	Aug	Sep	Oct
110	2.833394	1.129163	1.162841
115	1.417978	0.617046	0.658904
120	0.062811	0.126728	0.176418

125	−1.237028	−0.343571	−0.286369
130	−2.485879	−0.795423	−0.731003

Next, find for each of the 15 calls values for d_2:

X	Aug	Sep	Oct
110	2.801988	1.042362	1.074632
115	1.386572	0.530245	0.570695
120	0.031405	0.039928	0.088208
125	−1.268433	−0.430371	−0.374578
130	−2.517284	−0.882222	−0.819212

Now, find $N(d_1)$ for each of the 15 calls:

X	Aug	Sep	Oct
110	0.997697	0.870585	0.877553
115	0.921901	0.731398	0.745021
120	0.525041	0.550422	0.570017
125	0.108038	0.365584	0.387298
130	0.006462	0.213184	0.232388

Next, determine $N(d_2)$ for each of the 15 calls:

X	Aug	Sep	Oct
110	0.997461	0.851378	0.858730
115	0.917214	0.702029	0.715897
120	0.512527	0.515925	0.535145
125	0.102322	0.333463	0.353987
130	0.005913	0.188828	0.206333

Now use $N(d_1)$ and $N(d_2)$ to value the calls and put−call parity to value the puts.
CALLS

X	Aug	Sep	Oct
110	10.165	11.494	11.942
115	5.305	7.616	8.030
120	1.593	4.586	4.930
125	0.193	2.488	2.741
130	0.008	1.211	1.375

PUTS

X	Aug	Sep	Oct
110	0.003	0.701	0.666
115	0.134	1.787	1.685
120	1.415	3.721	3.537
125	5.009	6.587	6.290
130	9.816	10.274	9.866

The options whose values are underlined are overvalued by the market; they should be sold. Other options are undervalued by the market; they should be purchased.

9. Zero. This is because the futures contract can be replicated with a long position in a call and a short position in a put with the same exercise terms. The gammas of the call and the put are the same. A long position in a futures contract is replicated with a single long position in a call and a single short position in a put. Thus, the gamma of the long call position offsets the gamma of the short put position.

10. **a.** Tanker Company calls expiring in three months with an exercise price equal to 25 are currently worth 6.5725. This value is based on $d_1 = 0.799405189$, $d_2 = 0.499405189$, $N(d_1) = 0.787972249$, and $N(d_2) = 0.691253018$.

b. Greeks for the call are computed as follows:
Delta = 0.787972249
Gamma = 0.032206303
Theta = −6.070521207
Vega = 4.347657115
Rho = 4.266663344

c. $N(d_2) = 0.691253018$

d. $p_0 = 1.261959101$

e. Delta = −0.212027751
Gamma = 0.032206303

Theta = −9.503750259
Vega = 4.347657115
Rho = −1.905697909

11. a. Tanker Company calls expiring in three months with an exercise price equal to 30 are currently worth 3.7442. This value is based on $d_1 = 0.019166667$, $d_2 = -0.10833333333$, $N(d_1) = 0.575998341$, and $N(d_2) = 0.456865641$.

b. Greeks for the call are computed as follows:
Delta = 0.575998341
Gamma = 0.043524258
Theta = −7.727401147
Vega = 5.875513
Rho = 3.383927737

c. $N(d_2) = 0.456865641$.

d. Portfolio settings are determined by solving the following:

$$\begin{bmatrix} delta_P \\ gamma_P \\ \#_S \end{bmatrix} = \begin{bmatrix} delta_S & delta_{C1} & delta_{C2} \\ gamma_S & gamma_{C1} & gamma_{C2} \\ \#_S & 0 & 0 \end{bmatrix} \begin{bmatrix} \#_S \\ \#_{C1} \\ \#_{C2} \end{bmatrix}$$

$$\mathbf{r} = \mathbf{Y} \#$$

$$\begin{bmatrix} delta_P \\ gamma_P \\ \#_S \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0.787972 & 0.575998 \\ 0 & 0.0322 & 0.0435 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \#_S \\ \#_{C1} \\ \#_{C2} \end{bmatrix}$$

$$\mathbf{r} = \mathbf{Y} \#$$

$$\begin{bmatrix} \#_S \\ \#_{C1} \\ \#_{C2} \end{bmatrix} = \begin{bmatrix} 1 \\ -2.7643 \\ 2.045477 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 2.7643 & -36.5826 & -2.7643 \\ -2.04548 & 50.04546 & 2.04548 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

$$\# = \mathbf{Y}^{-1} \mathbf{r}$$

Thus, for each underlying share of stock purchased, a delta-gamma neutral portfolio requires a short position in 2.7643 calls with $25 exercise prices and a long position in 2.045477 calls with $30 exercise prices.

e. First, investment amounts in the stock and each of the options are computed as follows:
$\#c_{25}$: $-2.7643 \times 6.5725 = -18.1684$
$\#c_{30}$: $2.045477 \times 3.7442 = 7.6586$
$\#S$: $1 \times 30 = 30$
Total investment: $(-2.7643 \times 6.5725) + (2.045477 \times 3.7442) + (1 \times 30) = 19.4903$
Investment weights are computed as follows:
$W_{c25} = -18.1684/19.4903 = -0.9322$
$W_{c30} = 7.6586/19.4903 = 0.39294$
$W_S = 30/19.4903 = 1.53923$
Thus, the portfolio theta is computed as follows:
$W_{c25} \times \theta_{c25} + W_{c30} \times \theta_{c30} + W_S \times \theta_S$
$\theta_p = -0.9322 \times -6.0705 + 0.39294 \times -7.7274 + 1.53923 \times 0 = 2.622$

12. To answer this question, we first calculate d_1:

$$d_1 = \frac{\ln\left(\frac{0.7}{0.65}\right) + \left(0.06 - 0.04 + \frac{0.3^2}{2}\right) \cdot 2}{0.3\sqrt{2}} = 0.2925$$

Next we calculate d_2:
$d_2 = d_1 - \sigma\sqrt{T} = 0.2925 - 0.4242 = -0.1317$
Next, find cumulative normal density functions (z-values) for d1 and d2:
$N(d_1) = N(0.2925) = 0.6151$
$N(d_2) = N(-0.1317) = 0.4476$
Finally, we value the call as follows:
$c_0 = (0.65 \times 0.8870)(0.6151) - [0.7 \times 0.9231] \times (0.4476) = \$0.1133$
We can evaluate a put for this European currency option series using Equation (4) as follows:
$p_0 = c_0 + Xe^{-r(d)T} - s_0e^{-r(f)T} = 0.1133 + 0.65e^{-0.06 \times 0.5} - 0.7e^{-0.04 \times 0.5} = 0.092$

13. Based on the Garman-Köhlagen Model for currency calls followed by the put-call parity relation, we value the 2-year put on CNY as follows:

$$d_1 = \frac{ln(\frac{0.15}{0.2}) + (.02 - .08 + \frac{1}{2}2^2) \times 2}{.2\sqrt{2}} = -1.3$$

$$d_2 = -1.3 - .2\sqrt{2} = -1.583$$

$$N(d_1) = .0968;\ N(d_2) = .0567$$

$$c_0 = \frac{0.15}{e^{.08 \times 2}} \times .0968 - \frac{.2}{e^{.02 \times 2}} \times .0567 = .00147$$

$$p_0 = .00147 + .2 \times e^{-.02 \times 2} - .15e^{-.06 \times 2} = .0658$$

14. The two relevant options market hedging strategies are the put hedge strategy and the conversion. These are described as follows:

Put Hedge
Strategy: Purchase a six month put option on ¥15,000,000 with an exercise price of \$0.009/¥ and a premium of \$150. Time zero brokerage costs total \$750 (15 contracts at \$50 per contract—pretty high, given the premiums involved). Thus the total time zero cash outlay is \$900. Expressed in terms of future value, the total cash outlay is \$922.50 since interest forgone on the sum of the premium and brokerage costs totals \$22.50.
Result: Receive one of the following in six months:
An unlimited potential maximum less the \$922.50 premium and brokerage fees. The dollar value of this strategy increases as the value of the dollar drops.
A minimum of \$135,000 less \$922.50 for a net of \$134,077.50. This minimum value to be received might be unacceptably low for some managers; however, there is upside cash flow potential.

Conversion or Call and Put Hedge

Strategy: Through the combination of short call and long put positions, total risk can be eliminated. Consider the writing of a call with an exercise price of \$0.009 expiring in six months along with the purchase of a put with the same terms. The time zero cash flows are summarized as follows:

Put premium	−\$150	Call premium	+\$1500
Put brokerage fee	−\$750	Call brokerage fee	−\$750
Net time zero cash flows	−\$150		

Result: The forgone interest on the net time zero outlay is \$3.75. If the six-month exchange rate is less than \$0.009/¥, the exchange rate of \$0.009/¥ is locked in by the put. If the exchange rate exceeds \$0.009/¥, the obligation incurred by the short position in the call is activated. Thus, the exchange rate of \$0.009/¥ is locked in no matter what the exchange rate is. The cash flows in six months are summarized as follows:

Put cash flows (¥15,000,000 × MAX[0.009 − $S_{1,0}$])

Call cash flows (¥15,000,000 × MIN[0.009 − $S_{1,0}$])

Total of option transactions:

(¥15,000,000 × (0.009 − S_1)) = \$135,000 − (¥15,000,000 × S_1)

Exchange of currency = (¥15,000,000 × S_1)

Time zero cash flows = −\$150

Interest on time zero flows = \$3.75

Total time one cash flows = \$134,846.25

CHAPTER 10

1. a., b. See following table.

Date	Investor Contributions	Investor Withdrawals	Fund Dividends	Assets	Number of Shares	Net Asset Value	r_t
2/1	1,000,000	0	0	1,000,000	20,000	50.00	N/A
2/2	0	0	0	1,030,000	20,000	51.50	3.00%
2/3	0	0	30,000	1,030,000	20,000	51.50	2.91%
2/4	0	0	0	980,000	20,000	49.00	−4.85%
2/5	0	49,500	0	940,500	19,000	49.50	1.02%

c.

$$\overline{r}_{g,p} = \sqrt[n]{\prod_{t=1}^{n}(1+r_t)} - 1 = \sqrt[4]{(1.03)(1.0291)(0.9515)(1.102)} - 1 = 0.004676$$

2. a. NAV and DIV data for Fund A are followed by return calculations:

Date	t	NAV_t	NAV_{t-1}	DIV_t	r_t	NOTES
June 30	2	50	0	0	—	First month
July 31	2	55	50	0	0.100	$(55 \div 50) - 1 = 0.10$
Aug 13	3	50	55	0	−0.091	$(50 \div 55) - 1 = -0.091$
Sep 30	4	54	50	0	0.080	$(54 \div 50) - 1 = 0.08$
Oct 31	5	47	54	2	−0.092	ex-$2 dividend; $[(47 + 2) \div 54] - 1 = -0.092$
Nov 30	6	51	47	0	0.081	$(51 \div 47) - 1 = 0.081$

b. Monthly returns for Fund B beginning with July are 0.60, −0.50, 0.50, −0.50, and 0.50.

c. Geometric mean returns are calculated for Funds A and B (including their dividends) as follows:

$$\begin{aligned}\overline{r}_{g,A} &= \sqrt[n]{\prod_{t=1}^{n}(1+r_t)} - 1 \\ &= \sqrt[5]{(1+0.10)(1-0.091)(1+0.08)(1-0.092)(1+0.081)} - 1 \\ &= 0.0117\end{aligned}$$

$$\begin{aligned}\overline{r}_{g,B} &= \sqrt[n]{\prod_{t=1}^{n}(1+r_t)} - 1 \\ &= \sqrt[5]{(1+0.60)(1-0.50)(1+0.50)(1-0.50)(1+0.50)} - 1 \\ &= -0.0208\end{aligned}$$

d. Fund internal rates of return from part c are time-weighted returns.

3. a. Average historical returns (in decimal format) are as follows:

$$\overline{R}_p = 0.062$$

$$\overline{R}_L = 0.106$$

$$\overline{R}_M = 0.098$$

b.

$$\sigma_P^2 = 0.000696$$

$$\sigma_L^2 = 0.008824$$

$$\sigma_M^2 = 0.001576$$

c. $COV[P,M] = [(0.04 - 0.062) \cdot (0.15 - 0.098) + (0.07 - 0.062) \cdot (0.10 - 0.098) + (0.11 - 0.062) \cdot (0.03 - 0.098) + (0.04 - 0.062) \cdot (0.12 - 0.098) + (0.05 - 0.062) \cdot (0.09 - 0.098)\ ] \div 5 = -0.000956$

$$\rho_{P,M} = \frac{COV[P,M]}{\sigma_P \sigma_M} = \frac{-0.000956}{0.0264 \cdot 0.039} = -0.912$$

d. $COV[L,M] = [\ (0.15 - 0.098) \cdot (0.19 - 0.106) + (0.10 - 0.098) \cdot (0.04 - 0.106) + (0.03 - 0.098) \cdot (-0.04 - 0.106) + (0.12 - 0.098) \cdot (0.21 - 0.106) + (0.09 - 0.098) \cdot (0.13 - 0.106)] \div 5 = 0.003252$

$$\rho_{L,M} = \frac{COV[L,M]}{\sigma_M \sigma_L} = \frac{-0.003252}{0.039 \cdot 0.094} = 0.872$$

e. Stock or fund beta is $COV[S,M]/VAR[M]$. The two fund betas are:

$$\beta_P = \frac{COV\{P,M\}}{VAR[M]} = \frac{-0.000956}{0.001576} = -0.61$$

$$\beta_L = \frac{COV\{L,M\}}{VAR[M]} = \frac{-0.003252}{0.001576} = 2.06$$

f. The Treynor Index for the market is $T_M = (0.08 - 0.03)/1 = 0.05$.
The Treynor Index for the Patterson Fund was $T_P = (0.10 - 0.03)/(-0.61) = -0.1148$.
The Treynor Index for the Liston Fund was $T_L = (0.14 - 0.03)/(2.06) = 0.0539$.

g. The Treynor Index for the market is always simply the market's risk premium. Thus, there is nothing particularly noteworthy here other than the standard by which we compare the performance of the two funds. The Treynor Index for the Patterson Fund is negative, which would normally compare unfavorably to that of the market Treynor Index. However, since the risk premium for Patterson is positive while its beta is negative, this negative risk premium actually indicates strong risk-adjusted performance. Thus, Patterson outperformed the market on a risk-adjusted basis. Liston also outperformed the market on a risk adjusted basis since its Treynor Index of 0.0539 exceeded that of the Liston Fund.

4. a. First, calculate risk premiums for both the fund and the market as follows:

t	R_p	R_m	r_f	$R_p - r_f$	$R_m - r_f$
1	0.35	0.28	0.05	0.3	0.23
2	0.04	0.05	0.05	−0.01	0
3	0.1	0.09	0.05	0.05	0.04
4	−0.01	−0.02	0.05	−0.06	−0.07
5	0.38	0.32	0.05	0.33	0.27
6	0.31	0.25	0.05	0.26	0.2
7	0.33	0.26	0.05	0.28	0.21

8	0.42	0.3	0.05	0.37	0.25
9	0.07	0.14	0.05	0.02	0.09
10	−0.02	−0.06	0.05	−0.07	−0.11
11	−0.07	−0.15	0.05	−0.12	−0.2
12	−0.11	−0.21	0.05	−0.16	−0.26
13	0.42	0.26	0.05	0.37	0.21
14	0.01	0.04	0.05	−0.04	−0.01
15	0.05	0.08	0.05	0	0.03
16	0.07	0.11	0.05	0.02	0.06
17	−0.01	−0.03	0.05	−0.06	−0.08
18	−0.12	−0.37	0.05	−0.17	−0.42
19	0.33	0.3	0.05	0.28	0.25
20	0.05	0.13	0.05	0	0.08

Next, run a simple ordinary least squares regression of $(R_p - r_f)$ on $(R_m - r_f)$. The fund's beta is the slope term in the regression and the fund's alpha is its vertical intercept:

SUMMARY OUTPUT					
Regression Statistics					
Multiple *R*		0.910030353			
R Square		0.828155243			
Adjusted *R* Square		0.818608312			
Standard Error		0.079179681			
Observations		20			
ANOVA					
	df	SS	MS	*F*	Significance *F*
Regression	1	0.543845407	0.543845407	86.74570346	2.63657E − 08
Residual	18	0.112849593	0.006269422		
Total	19	0.656695			
	Coefficients	Standard Error	*t* Stat	*P*-value	
Intercept	0.045004834	0.018088349	2.488056431	0.022868369	
X Variable 1	0.895978331	0.096199656	9.313737352	2.63657E − 08	

Thus, the fund beta is 0.8959.

b. The fund alpha is 0.04. Since the alpha is statistically significant at the 5% level, we can conclude that the fund has outperformed the market.

c. Risk premiums and relevant squared values are given as:

t	$R_p - r_f$	$R_m - r_f$	$(R_m - r_f)^2$	t	$R_p - r_f$	$R_m - r_f$	$(R_m - r_f)^2$
1	0.3	0.23	0.0529	11	−0.12	−0.2	0.0400
2	−0.01	0	0.0000	12	−0.16	−0.26	0.0676
3	0.05	0.04	0.0016	13	0.37	0.21	0.0441
4	−0.06	−0.07	0.0049	14	−0.04	−0.01	0.0001
5	0.33	0.27	0.0729	15	0	0.03	0.0009
6	0.26	0.2	0.0400	16	0.02	0.06	0.0036
7	0.28	0.21	0.0441	17	−0.06	−0.08	0.0064
8	0.37	0.25	0.0625	18	−0.17	−0.42	0.1764
9	0.02	0.09	0.0081	19	0.28	0.25	0.0625
10	−0.07	−0.11	0.0121	20	0	0.08	0.0064

Regression results are given as follows:

SUMMARY OUTPUT					
Regression Statistics					
Multiple *R*	0.910030353				
R Square	0.828155243				
Adjusted *R* Square	0.818608312				
Standard Error	0.079179681				
Observations	20				
ANOVA					
	df	SS	MS	*F*	Significance *F*
Regression	1	0.543845407	0.543845407	86.74570346	2.63657E − 08
Residual	18	0.112849593	0.006269422		
Total	19	0.656695			
	Coefficients	Standard Error	*t* Stat	*P*-value	
Intercept	0.045004834	0.018088349	2.488056431	0.022868369	
X Variable 1	0.895978331	0.096199656	9.313737352	2.63657E − 08	

Since the regression coefficient for the squared market risk premium is positive and statistically significant at the 1% level, the Ripco Fund does present evidence of market timing ability.

d. Risk premiums and interaction terms are given as follows:

t	$R_p - r_f$	$R_m - r_f$	$D(R_m - r_f)$	t	$R_p - r_f$	$R_m - r_f$	$D(R_m - r_f)$
1	0.3	0.23	0	11	−0.12	−0.2	−0.2
2	−0.01	0	0	12	−0.16	−0.26	−0.26
3	0.05	0.04	0	13	0.37	0.21	0
4	−0.06	−0.07	−0.07	14	−0.04	−0.01	−0.01
5	0.33	0.27	0	15	0	0.03	0
6	0.26	0.2	0	16	0.02	0.06	0
7	0.28	0.21	0	17	−0.06	−0.08	−0.08
8	0.37	0.25	0	18	−0.17	−0.42	−0.42
9	0.02	0.09	0	19	0.28	0.25	0
10	−0.07	−0.11	−0.11	20	0	0.08	0

Regression results are given as follows:

SUMMARY OUTPUT					
Regression Statistics					
Multiple *R*		0.974705741			
R Square		0.950051281			
Adjusted *R* Square		0.944174961			
Standard Error		0.043925789			
Observations		20			
ANOVA					
	df	SS	MS	*F*	Significance *F*
Regression	2	0.623893926	0.311946963	161.6745344	8.65879E − 12
Residual	17	0.032801074	0.001929475		
Total	19	0.656695			
	Coefficients	Standard Error	*t* Stat	*P*-value	
Intercept	−0.014398622	0.013629095	−1.05646213	0.305539994	
X Variable 1	0.984588481	0.055112431	17.86508898	1.87737E − 12	
X Variable 2	1.58370714	0.245876938	6.441056058	6.09187E − 06	

Since the regression coefficient for the interaction term is positive and statistically significant at the 1% level, the Ripco Fund does present evidence of market timing ability.

5. The Sharpe ratio adjusts return performance for standard deviation, which captures both systematic and unsystematic risk. It is less appropriate for the evaluation of the component of a portfolio because it captures unsystematic risk, which can be diversified away in a more efficient portfolio. Similarly, the Treynor and Jensen measures do not account for unsystematic risk, which can be diversified away, and are therefore more appropriate for portfolio components or sectors.
6. The daily VWAP for stock X is calculated as follows:

$$\begin{aligned} VWAP = \frac{\sum_j Q_j P_j}{\sum_j Q_j} &= (30,000 \times 100.00 + 10,000 \times 100.00 + 40,000 \times 100.02 + 10,000 \\ &\quad \times 100.01 + 10,000 \times 100.05 + 10,000 \times 100.03 + 20,000 \times 100.06 \\ &\quad + 20,000 \times 100.04 + 30,000 \times 100.07 + 40,000 \times 10.05) + 10,000 \\ &\quad \times 100.09 + 40,000 \times 100.05)/(30,000 + 10,000 + 40,000 + 10,000 \\ &\quad + 10,000 + 10,000 + 20,000 + 20,000 + 30,000 + 40,000 \\ &\quad + 10,000 + 40,000 = 27,010,700/270,000 \\ &\quad = 100.0396296 \end{aligned}$$

The VWAP excluding the broker's transactions were:

$$\begin{aligned} VWAP = \frac{\sum_j Q_j P_j}{\sum_j Q_j} &= (30,000 \times 100.00 + 40,000 \times 100.02 + 10,000 \times 100.05 + 20,000 \\ &\quad \times 100.06 + 30,000 \times 100.07 + 10,000 \times 100.09)/(30,000 + 40,000 \\ &\quad + 10,000 + 20,000 + 30,000 + 10,000) \\ &\quad = 100.0392857 \end{aligned}$$

Since $100.0396296 > 100.0392857$, meaning that the average price paid in the market with the broker's transactions were higher, the broker was outperformed (slightly) by the market. The broker's VWAP without the rest of the market was 100.0396, higher than that of the market.

7. **a.** Factors that affect implementation shortfall are those that impact a trader's ability to execute an order at or close to the decision price. These factors include the broker's commission, the size of the order, market liquidity for the security, and the volatility of the stock, which itself, might be affected by correlations in prices from one trade to the next.
   b. Increasing the speed at which a large transaction is executed exerts increased price pressure on the security, increasing slippage and implementation shortfall.

c. Decreasing the speed at which a large transaction is executed may increase differences between decision and execution prices. Delays in execution potentially increase delay costs and potentially increase opportunity costs in the event that the transaction is not fully executed. These effects are exacerbated in a more volatile market. If the speed of execution is reduced by slicing the order into small transactions, total brokerage commissions might be higher.

8. a. The ideal portfolio profit is \$200 = 1000(\$20.20 − \$20.00). The actual portfolio profit is \$28 = 400(\$20.20 − \$20.10) + 400(\$20.20 − \$20.15) − \$32. The difference, implementation shortfall is \$172. Note that these calculations can be realized summing component costs as in part b.

b. Commissions are \$32, delay costs are \$100, price impact costs are \$20, and opportunity costs are \$20. Notice that these costs sum to \$172. However, one might be able to argue that some portion of the delay costs should actually be price impact costs given the size of the limit order placed at the open.

c. Yes: Favorable swings in the price of the security (decreased prices for purchasing and increased prices for sales) can reduce, eliminate, or even turn negative implementation shortfall.

9. First, note that normal curve areas (z-values from Table A.1 in Mathematics Appendix A.4) consistent with these probabilities are:

$$0.95\text{: } 1.645$$
$$0.99\text{: } 2.326$$

VaR for each of the 4 circumstances is calculated as follows:

$$VaR = \$80{,}000{,}000 \times 0.01 \times z \times \sqrt{\text{t}}$$
$$VaR = \$80{,}000{,}000 \times 0.01 \times 1.645 \times \sqrt{1} = 1{,}316{,}000$$
$$VaR = \$80{,}000{,}000 \times 0.01 \times 2.326 \times \sqrt{1} = 1{,}860{,}800$$
$$VaR = \$80{,}000{,}000 \times 0.01 \times 1.645 \times \sqrt{30} = 7{,}208{,}028.86$$
$$VaR = \$80{,}000{,}000 \times 0.01 \times 2.326 \times \sqrt{30} = 10{,}192{,}021.35$$

Thus, we can infer with 95% certainty that the portfolio's loss on a given day will not exceed \$1,316,000. Similarly, we can infer with 99% certainty that the portfolio's loss on a given day will not exceed \$1,860,800. We can also infer with 95% and 99% certainty that the portfolio's loss on in a given month will not exceed \$7,208,028.86 and \$10,192,021.35.

10. We will start by checking a *z*-table or spreadsheet to find the one-tail value that corresponds to a 95% confidence interval. We find that $z = 1.645$ (corresponds to *z*-value of .95 (or 1 − 5%). Now, we can calculate *VaR*:

$$VaR = \$100,000 \times .01 \times 1.645 \times \sqrt{4} = \$3,290$$

Assuming that daily asset returns are normally distributed (often a questionable assumption), uncorrelated over time and with a standard deviation of .01, there is a 5% probability that you will experience a loss exceeding \$3,290 during any given 4-week month. The one-tailed *z*-value corresponding with the 5% confidence interval is 1.645. Thus, 95% of the time, your losses will be less than the computed *VaR* figure of \$3,290.

11. First, perhaps the most significant criticism of *VaR* methodology concerns the assumption of normally distributed returns, suggesting that the most disastrous outcomes (sometimes referred to in a risk measurement context as black swans) are very unlikely. If worst-case or near-worst-case scenarios are not actually rare, those situations may be parts of "fat tails," and their probabilities might be underestimated by the thin tails consistent with a normal distribution. Remodel with another distribution if necessary. Second, if standard deviation is not the best or most complete measure of variability of returns, or if variability of returns does not provide a good indicator of risk, then adapt the *VaR* method to a more appropriate measure of risk, and/or rely on other risk measurement methods.

12. **a.** Security weights are: $w_X = .16667$, $w_Y = .33333$, $w_Z = .5$

$$E[R_p] = (.16667 \times .10) + (.33333 \times .15) + (.5 \times .20) = .167$$

$$\sigma_p^2 = (.16667 \times .16667 \times .12 \times .12 \times 1) + (.16667 \times .33333 \times .12 \times .18 \times .8)$$

$$+ (.16667 \times .5 \times .12 \times .24 \times .7) + (.33333 \times .16667 \times .18 \times .12 \times .8)$$

$$+ (.33333 \times .33333 \times .18 \times .18 \times 1) + (.33333 \times .5 \times .18 \times .24 \times .6)$$

$$+ (.5 \times .16667 \times .24 \times .12 \times .7) + (.5 \times .33333 \times .24 \times .18 \times .6)$$

$$+ (.5 \times .5 \times .24 \times .24 \times 1) = .03232; \sigma_p = .1798$$

b. $VaR = \$6{,}000 \times .1798 \times \sqrt{52} \times 2.326 = \$347.98$, which implies that the maximum loss incurred on this portfolio over the course of one week is $347.98 with a confidence level of 99%.

CHAPTER 11

1. **a.** $E[U_{w,\ No\ Gamble}] = \ln(2) = 0.693147$

b. Solve the following for G, where G is the cost of the gamble and x is its winnings:

$$E[U_{w,\ With\ gamble}] = \sum_{i=1}^{\infty} p_i U(w + x_i - G) = \sum_{i=1}^{\infty} [0.5^i \ln(2 + 2^i - G)] = 0.693147$$

$G = 3.34757$

Note: A spreadsheet may be useful to solve this infinite series. The value G is the payment for the gamble, the initial wealth level is 2, winnings are 2^i where i is the number of tosses before the first head. The value of G is obtained by iteration. The following is the first 13 rows of spreadsheet calculations for this problem:

	p^i	$(2 - G + 2^i)$	$(p^i)(2 - G + 2^i)$	SUM$(p^i)(2 - G + 2^i)$
1	0.5	−0.42705	−0.213525713	−0.213525713
2	0.25	0.975476	0.24386905	0.030343338

3	0.125	1.894982	0.236872775	0.267216113
4	0.0625	2.684606	0.167787887	0.435004
5	0.03125	3.422712	0.106959748	0.541963748
6	0.015625	4.137602	0.064650039	0.606613787
7	0.0078125	4.841447	0.037823801	0.644437588
8	0.00390625	5.5399	0.021640233	0.666077821
9	0.00195313	6.235689	0.01217908	0.678256901
10	0.00097656	6.930155	0.006767729	0.68502463
11	0.00048828	7.623961	0.003722637	0.688747268
12	0.00024414	8.317437	0.002030624	0.690777892
13	0.00012207	9.010749	0.001099945	0.691877837

A trial value G is entered elsewhere in the spreadsheet and this cell is referenced for all other cells where G is used. The value for G is iterated until the sum is sufficiently close to the natural log of 2. In this table, the value 3.34757 is used, where this value was obtained by trial and error in an effort to obtain 0.693147 (or some sufficiently close value) for the sum in the 13th row.

c. First, find the utility of \$1000: ln(1000) = 6.907755.
Now, solve the following for G, where G is the cost of the gamble:

$$E\left[U_{w,\ With\ gamble}\right] = \sum_{i=1}^{\infty} p_i U(w + x_i - G) = \sum_{i=1}^{\infty} \left[0.5^i \ln(2 + 2^i - G)\right] = 0.6907755$$

$G = 10.954$

Note: See the following table excerpted from a spreadsheet used to solve the infinite series, iterating for G. The key column, "Contribution to Utility", equals Probability $\times$ ln(1000 + "Gamble Payoff" $-$ G). The column is then summed such that the sum equals the utility of \$1000:

Toss	Probability	Gamble Payoff	Log of Post-Gamble Wealth	Contribution to Utility	Contribution to Wealth
1	0.5	2	6.898761	3.449380476	1
2	0.25	4	6.900777	1.725194247	1
3	0.125	8	6.904797	0.863099613	1
4	0.0625	16	6.912789	0.432049287	1
5	0.03125	32	6.928583	0.216518215	1
6	0.015625	64	6.959442	0.108741284	1

7	0.0078125	128	7.018443	0.054831586	1
8	0.00390625	256	7.126928	0.027839562	1
9	0.00195313	512	7.313917	0.014284995	1
10	0.00097656	1024	7.607404	0.007429106	1
11	0.00048828	2048	8.018641	0.003915352	1
12	0.00024414	4096	8.534059	0.002083511	1
13	0.00012207	8192	9.124896	0.001113879	1
14	6.1035E − 05	16384	9.762675	0.000595866	1
15	3.0518E − 05	32768	10.42694	0.000318205	1
16	1.5259E − 05	65536	11.10533	0.000169454	1
17	7.6294E − 06	131072	11.79102	8.99583E − 05	1

d. Solve the following for G, where G is the cost of the gamble:

$$E\left[U_{w,\ With\ \ gamble}\right] = \sum_{i=1}^{\infty} p_i U(w + x_i - G) = \sum_{i=1}^{\infty} \left[0.5^i \ln(2 + 2^{2i-1} - G)\right]$$
$$= (2 + 1 - G) + (2 + 8 - G) + (2 + 32 - G) + (2 + 64 - G) + \cdots$$

For any finite value of G, expected utility must equal ∞. Thus $G = \infty$, and an investor would be willing to pay any finite sum for this gamble. That is, this illustration shows how an investor, with diminishing marginal utility of wealth (log utility function) and risk aversion would still be willing to pay an infinite sum of money for a gamble. This example might be referred to as a "Super St. Petersburg Paradox."

2. In a pairwise choice between ice cream and cake, ice cream would be preferred by two out of three traders. In a pairwise choice between ice cream and cookies, cookies would be preferred by two out of three traders. In a pairwise choice between cake and cookies, cake would be preferred by two out of three traders. Thus, the social preferences, as indicated by the votes taken by traders, would be cookies $\succ$ ice cream $\succ$ cake $\succ$ cookies. Obviously, the social preferences, as determined by majority vote, are not transitive (rational).

3. **a.** Based on expected value, the actuarial value of this policy is $0.05 \times \$20{,}000 = \$1000$.
b. $\$1200 - 1000 = \$200$
c. $\$1000 - 1200 = -\$200$; $200 expected loss to the consumer
d. The sale is a rational transaction if the insurance company intends to increase its wealth (more is preferred to less).
e. The purchase is a rational transaction to the consumer if she is sufficiently risk-averse.

4. The two sets of choices are identical in terms of terminal wealth to the subject. However the phrasing of the question causes the choice sets to invoke different emotional responses; that is, the framing is different. This leads to the framing versus

substance problem, which is a violation of the von Neumann-Morgenstern independence axiom.

5. **a.** There should be no statistically significant difference
   b. Selling prices will exceed purchase prices. Kahneman and Tversky actually performed this experiment and found that the median selling price was $5.79 and the median purchase price was $2.25, a ratio of more than 2. On the possibility that this result might have been due to "wealth effects"—subjects given mugs were simply wealthier than those not given mugs, and this might drive the price differences. But the experiment was repeated where selling prices of one group were compared to the "choosing" prices of the other. In this scenario, subjects selected from listings of money amounts whether they would prefer to have a mug or money. Choosers are in precisely the same wealth position as sellers—they choose between a mug and money. The only difference is that sellers are "giving up" a mug they "own," whereas choosers are merely giving up the right to have a mug. The results were consistent. The median choosing price was half the median selling price ($3.50 vs. $7.00). Similar experiments have been performed in many markets with consistent results.
   c. Selling prices are likely to exceed buying prices by larger amounts. Investors will be more reluctant to sell stock that they already own and would be reluctant to purchase shares.
6. Option sets i and ii establish that people are impatient and associate a time value with money and good meals. However, this impatience and positive time value is inconsistent with the preference revealed in decision set iii. Apparently, in this scenario, people look forward to improved meals, as though they would give up time value to see an upward trend. This experiment was conducted by Lowenstein and Prelec (1993).
7. Each contestant has his own method for estimating the number of jelly beans, based on his abilities, experiences, visual and quantitative skills, etc. Each contestant, who possesses each of these and other attributes to at least a degree, incorporates all of these attributes into his best estimate. Since each contestant makes every effort to incorporate all of these attributes into the estimation process, there will be at least some systematic and correlated individual strengths in the estimation process of the population or consensus. However, each contestant also incorporates his weaknesses and biases in the estimation process, despite his best efforts not to. If these weaknesses in the estimation process are uncorrelated, and then consensus estimates will be unbiased.
8. First, since a 0.66 probability of a $2400 payout is being shifted to 0 from A and B to A^* and B^*, we will rewrite the statement of gamble payoffs as follows:

   Gamble A: 0.33 probability of receiving 2500, 0.66 of receiving 2400, and 0.01 of receiving 0

   Gamble B: 0.34 probability of receiving 2400 and 0.66 of receiving 2400

   and

   Gamble $A*$: 0.33 probability of receiving 2500, 0.01 of receiving 0, and 0.66 of receiving 0

Gamble $B*$: 0.34 probability of receiving 2400 and 0.66 of receiving 0

The investor is indifferent between Gambles A and B. Recall that the strong independence axiom states that if $x_j \succ x_k$, and then for any $\alpha \in [0,1]$, $\alpha x_i + (1 - \alpha)x_k \sim \alpha x_j + (1 - \alpha)x_k$. This strong independence axiom implies that for any $\alpha \in [0,1]$:

$$\begin{gathered}\alpha(0.33 \text{ prob. of receiving 2500 and 0.01 of receiving 0}) \\ +(1-\alpha)(0.66 \text{ prob. of receving 2400}) \sim \alpha(0.34 \text{ prob. of receiving 2400}) \\ +(1-\alpha)(0.66 \text{ prob. of receving 2400})\end{gathered}$$

which implies that:

$$(0.33 \text{ prob. of receiving 2500 and 0.01 of receiving 0}) \sim (0.34 \text{ prob. of receving 2400})$$

The same decomposition for Gambles A* and B* results in:

$$\begin{gathered}\text{Gamble } A*: \alpha(0.33 \text{ prob. of receiving 2500 0.01 of receiving 0}) \\ +(1-\alpha)(0.66 \text{ prob. of receiving 0})\end{gathered}$$

$$\text{Gamble } B*: \alpha(0.34 \text{ prob. of receiving 2400}) + (1-\alpha)(0.66 \text{ prob. of receiving 0})$$

which, by the same strong independence axiom, reduces to a comparison between:

Gamble $A*$: (0.33 prob. of receiving 2500 and 0.01 of receiving 0)

Gamble $B*$: (0.34 prob. of receiving 2400)

We know from our statement above concerning Gambles A and B, (0.33 prob. of receiving 2500 and 0.01 of receiving 0) $\sim$ (0.34 prob. of receiving 2400), that the investor must be indifferent between Gambles A* and B*.

9. a. This is a simple example of the framing problem discussed by Kahneman and Tversky.

b. The first set of choices is framed in terms of lives saved, or gains. The second set is framed in terms of lives lost, or losses. Framing the problem in terms of lives saved triggers risk aversion, while framing the same problem in terms of lives lost triggers risk seeking. Again, this is consistent with Kahneman and Tversky's prospect theory.

10. a. $0.5(300{,}000 - 100{,}000) + 0.5(0 - 100{,}000) = 50{,}000$

b. $[0.5(200{,}000 - 50{,}000)2 + 0.5(-100{,}000 - 50{,}000)2].5 = 150{,}000$

c. $5 \times 50{,}000 = 250{,}000$

Alternatively, in five wagers, with 120 (5!) possible win/loss scenarios, there is one scenario in which five consecutive losses totaling 500,000 occur with a probability of 0.5^5, five scenarios in which four losses and one win occur with a net loss of 200,000, and a probability of 5×0.5^5 and so on:

$1 \times 0.5^5 \times (-500{,}000) + 5 \times 0.5^5 \times (-200{,}000) + 10 \times 0.5^5 \times 100{,}000 + 10 \times 0.5^5 \times 400{,}000 + 5 \times 0.5^5 \times 700000 + 1 \times 0.5^5 \times 1{,}000{,}000 = 250{,}000$

d. $(1 \times 0.5^5 \times (-500{,}000 - 250{,}000)^2 + 5 \times 0.5^5 \times (-200{,}000 - 250{,}000)^2 + 10 \times 0.5^5 \times (100{,}000 - 250{,}000)^2 + 10 \times 0.5^5 \times (400{,}000 - 250{,}000)^2 + 5 \times 0.5^5 \times (700{,}000 - 250{,}000)^2 + 1 \times 0.5^5 \times (1{,}000{,}000 - 250{,}000)^2)^{0.5} = 335{,}410.2$

Alternatively, since payoffs from each of the five wagers are independent of one another, this standard deviation of five gambles can be computed as follows:

$$\left(5 \times 150,000^2\right)^{.5} = 335,410.2$$

e. Obviously, the set of five wagers described in parts c and d has the higher expected value.

f. Obviously, the set of five wagers described in parts c and d has the higher standard deviation. However, note that this higher standard deviation is less than five times the individual wager standard deviation.

g. The answer to this depends on how you evaluate the wagers and your own individual preferences. However, the reward to risk ratio for the single wager 50,000/150,000 = 3 is much greater than the reward to risk ratio of the set of 5 gambles 250,000/335.410.2 = 0.746. This suggests that as the number of gambles increases, the reward to risk ratio will also increase due to diversification. However, consider the response to the next part of this question.

h. Suppose that you find the individual wager described in parts a and b unacceptable. Then, if after having wagered four times (the first four of five wagers described in parts c and d), you have the opportunity to wager a fifth time, you should decline, since you find any single wager of this type to be unacceptable. Then, by the same logic, after having wagered three times, you would find the fourth wager unacceptable, and so on. Thus, any person finding the single wager described in parts a and b to be unacceptable should also find the set of five wagers to be unacceptable. Hence, diversification over time or over a series of sequential gambles would not mitigate risk. This is the substance of Paul Samuelson's "law of large numbers" fallacy (Samuelson, 1963).

CHAPTER 12

1. New information arrives randomly; otherwise it is not news. Thus, returns will fluctuate randomly as prices react to the random arrival of news.

2. Three-day moving averages are computed as follows:

t	P_t	MA_t
1	50	NA
2	51	NA
3	52	51
4	58	53.67
5	56	55.33

Our three-day moving average would suggest that investors sell the stock on each of days 3, 4, and 5.

3. The five-day simple weighted average is computed as follows:

NA	NA	NA	NA	6	7	7.6	9	10.8	11.6	12.8	14.6

4. No: Seasonal cycles are anticipated by shareholders who anticipate and discount cash flows accordingly. However, unanticipated changes in sales revenue would affect share prices.

5. Ten spin-offs provides for a very small sample set. Did stock returns increase for the year? Since returns were collected for only one year, and that year may not have both longer periods of decline and longer periods of increase, the returns might simply have been normal market returns. This scenario is especially likely if stock returns have not been adjusted for normal market returns.

6. Either service can be used equally well to time market decisions.

7. **a.** Acquiring company daily stock returns, $(P_t/P_{t-1}) - 1$, are computed as follows:

Day	Company X Return	Company Y Return	Company Z Return
−6	NA	NA	NA
−5	0	0	0.00207
−4	0.00249	0.00625	−0.00413
−3	0	0.00621	−0.00207
−2	0.00248	0.00617	−0.00207
−1	−0.00248	0	0.00208
0	0.03980	0.04907	0.00831
1	0.00239	−0.00584	0.00206
2	−0.00238	0.00588	0
3	0.00239	0.00584	0.00205
4	0.00238	−0.00581	0
5	−0.00238	0.00584	0

b. Excess returns (residuals) are computed in the following table:

Day	Company X Residual	Company Y Residual	Company Z Residual	Average Residual
−5	−0.000465	−0.000520	0.001988	0.000333
−4	0.002028	0.005729	−0.004214	0.001181
−3	−0.000465	0.005690	−0.002156	0.001022
−2	0.002021	0.005652	−0.002161	0.001837

−1	−0.002947	−0.000520	0.002001	−0.000488
0	0.039335	0.048559	0.008233	0.032042
1	0.001926	−0.006368	0.001979	−0.000820
2	−0.002852	0.005361	−0.000082	0.000809
3	0.001926	0.005327	0.001975	0.003076
4	0.001920	−0.006334	−0.000082	−0.001498
5	−0.002846	0.005327	−0.000082	0.000799

c. Average residuals for the stocks on each date along with standard deviations and normal deviates are computed as follows from Part b:

Day	Average Residual	Standard Deviation	Normal Deviate
−5	0.000333	0.00143	0.23248
−4	0.001181	0.00503	0.23501
−3	0.001022	0.00413	0.2475
−2	0.001837	0.00391	0.46985
−1	−0.000488	0.00247	−0.1972
0	0.032042	0.02113	1.5165
1	−0.00082	0.0048	−0.1707
2	0.000809	0.00418	0.19362
3	0.003076	0.00195	1.57778
4	−0.001498	0.00431	−0.3479
5	0.000799	0.00416	0.1922

We shall assume the residuals follow a t-distribution and we will perform a one-tailed test with a 95% level of significance. Given $1 = 3 - 2$ degrees of freedom, the critical value for each test will be 6.314. Based on our computations above, we find that none of the residual t-statistics (normal deviates) exceed 6.314. Thus, we may not conclude with a 95% level of confidence that any residual differs from zero.

d. Cumulative average residuals are based on the following formula:

$$CAR_t = \sum_{i}^{t} AR_i$$

e. Cumulative average residuals are computed in our example as in either of the two following tables:

Day	Average Residual	Cumulative Average Residual
−5	0.000333	0.000333
−4	0.001181	0.001514
−3	0.001022	0.002537
−2	0.001837	0.004375
−1	−0.000488	0.003886
0	0.032042	0.035929
1	−0.000820	0.035108
2	0.000809	0.035917
3	0.003076	0.038993
4	−0.001498	0.037495
5	0.000799	0.038294

Day	Cumulative Residual, X	Cumulative Residual, Y	Cumulative Residual, Z	Cumulative Average Residual
−6	NA	NA	NA	NA
−5	−0.00046	−0.00052	0.00198	0.000333
−4	0.00156	0.00520	−0.00226	0.001514
−3	0.00109	0.01089	−0.00438	0.002537
−2	0.00311	0.01655	−0.00654	0.004375
−1	0.00017	0.01603	−0.00454	0.003886
0	0.03950	0.06459	0.00369	0.035929
1	0.04143	0.05822	0.00567	0.035108
2	0.03858	0.06358	0.00558	0.035917
3	0.04050	0.06891	0.00756	0.038993
4	0.04245	0.06257	0.00748	0.037495
5	0.03958	0.06790	0.00739	0.038294

f. Standard deviations and normal deviates are computed in the following table:

Day	Standard Deviation	Normal Deviate
−5	0.001432	0.23307
−4	0.003717	0.40749
−3	0.007742	0.32775
−2	0.011599	0.37720
−1	0.010778	0.36056
0	0.012468	3.77089
1	0.021347	1.66432
2	0.023212	1.56549
3	0.024538	1.60622
4	0.022204	1.70760
5	0.024190	1.60047

g. Normal deviates do not exceed the critical value of 6.314. Thus, if our hypotheses concerning each date t in our testing period were given as follows:
H_0: $CAR_t \leq 0$ H_A: $CAR_t > 0$,
We would not be able to reject the null hypothesis that $CAR_t >$ with 95%confidence for any date.

8. a. Average residuals, standard deviations, and normal deviates are given in the following table:

Average Residuals

AR_t	σ_t	Normal Deviate
0.006628066	0.010006	0.662407616
0.009156222	0.023227	0.394210557
0.025319415	0.023516	1.076709668
0.095348093	0.017464	5.459552045
0.006292193	0.025738	0.244470407
0.003757852	0.018872	0.199125429
0.000728706	0.02107	0.03458566

b. Cumulative average residuals for each stock follow:

Cumulative Average Residuals for Individual Stocks

Day	Stock 1	Stock 2	Stock 3	Stock 4	Stock 5	Stock 6	Stock 7	Stock 8	Stock 9	Stock 10
−3	0.003279	−0.00814	0.008945	−0.00255	0.011395	−0.00797	0.011223	0.014037	0.020344	0.015708
−2	0.007719	−0.0281	0.026009	0.001244	0.059814	0.009998	0.000247	−0.01064	0.044042	0.047514
−1	0.03971	−0.00554	0.012505	0.027555	0.1298	0.05721	0.012878	0.027403	0.059544	0.049967
0	0.131334	0.080815	0.10067	0.105526	0.258151	0.169602	0.082141	0.116799	0.167306	0.152173
1	0.174439	0.109783	0.128434	0.112295	0.209142	0.158994	0.092843	0.107061	0.172755	0.161694
2	0.181154	0.121328	0.142789	0.125576	0.182949	0.125532	0.119173	0.110469	0.188128	0.167919
3	0.16173	0.116801	0.135713	0.124435	0.178023	0.154783	0.111341	0.077968	0.213058	0.198455

Cumulative average residuals, standard deviations, and normal deviates follow:

Cumulative Average Residual Statistics

Day	CAR_t	σ_t	Normal Deviate
−3	0.006628	0.010006	0.662408
−2	0.015784	0.027937	0.565002
−1	0.041104	0.037577	1.09386
0	0.136452	0.053544	2.548386
1	0.142744	0.037887	3.76762
2	0.146502	0.030352	4.826832
3	0.147231	0.041947	3.509878

c. Average residuals are statistically significant at the 1% level only on day 4, the event day. Cumulative average residuals are statistically significant at the 1% level on days 5 through 7, and at the 5% level on day 4.

9. The Hollywood Stock Exchange can provide information about the potential success for the movie, where that success will likely impact the film company's stock price and be of interest to prospective investors. These prices may also provide useful information on other film company offerings as well as trends useful to the film industry. Such information will also be useful to managers of firms in the industry, helping them to determine and plan movie ideas to explore and movies to produce and to distribute.

10. a. Fairly quickly after the timely arrival of the mail ship, English security prices would change in the same direction and by a similar proportion as they did in London markets. Thus, Amsterdam price shifts would follow London price shifts three days earlier.

b. Fairly quickly after the delayed arrival of the mail ship, English security prices would change in the same direction and by a similar proportion as they did in London markets 3 + 4 days earlier (see Koudijs [2015], Figure 4), though with additional price volatility.

CHAPTER 13

1. Consider the following, based on Griffin and Sham [2017]: If prices deviate from fundamentals, traders have an economic incentive to trade against the deviations. Trading to exploit deviations from fundamentals is facilitated by high market liquidity, low transactions costs, arbitrage possibilities, etc., which are characteristic of well-developed and liquid markets. If a manipulator successfully causes a deviation of a price from its fundamental value, the price will typically revert to or nearer to its fundamental value when the trader exits her position. Thus, while price slippage may aid the manipulator when initiating the manipulating transactions, slippage works against the manipulator on exit. This diminishes the profit potential of market manipulation through manipulator trading activity. In addition, a manipulator must produce enough price discrepancy from the security's fundamental value to cover the expected penalty costs of security fraud. These penalty costs are enhanced by an effective regulatory system, a characteristic of the typical well-developed market.

2. These are really just rhetorical questions. Nonetheless, consider the following:

a.
- Inside information in stock markets is often considered to be the property of the firm's owners. This is a view frequently taken by the U.S. courts (see Chapter 5). Trading on inside information can violate the insider's fiduciary duties that the firm's officers and directors owe to their shareholders. No such fiduciary duties are violated in real estate and other markets, nor is there a violation of a relationship of trust and confidence.
- Inside traders in real estate and other markets do not obtain their inside information through illegal means or the breach of professional ethics.
- Because officers and directors control the production and disclosure of inside information, they can use this information to transfer wealth from shareholders to themselves. The sellers of real estate and other assets are the beneficiaries of their own inside information.
- The primary economic rationale advanced for prohibiting insider trading is that such trading can adversely affect securities markets. Prohibiting insider trading in labor (if it is possible) and real estate markets would probably destroy those markets.

b. First, information, especially inside information is a valuable commodity. Permitting insider trading would encourage increased production of inside information. Second, insider trading is likely to improve price discovery. Third, long-term (buy and hold) investors are not likely to be hurt by insider trading as they don't sell.

3. Insider trading does not violate the Securities Act of 1933. It violates the Securities and Exchange Act of 1934, which deals with secondary markets for securities.

4. **a.** Shenker was accused of *spoofing*, placing buy and sell orders that he had no intent of executing. These orders were placed for the purpose of manipulating the prices of the affected securities to obtain better executions on larger orders. Actually, this case was related to the Pomper case discussed in the text section on spoofing.
   b. The actual transactions were not used to manipulate the market; only quotes that he intended to cancel were used to manipulate the market.
   c. Any attempt to manipulate securities prices in secondary markets is illegal. Of course, spoofing specifically is banned by Dodd-Frank.
5. **a.** First, there could be a requirement that fees paid and rebates received be passed from the broker through to the client. This should eliminate the broker's conflict of interest. Second, the 0.3 cents per share limit imposed by Rule 610 of Reg NMS could be reduced or eliminated, diluting or eliminating maker-taker pricing, but potentially impairing liquidity. Finally, maker-taker fees could be reflected in the quotes displayed in markets, though this would probably necessitate allowing quotes in sub-penny increments.
   b. First, order cancellations might simply be banned, but this would likely cause more trading problems than it would address, because it is often necessary for traders to cancel bona-fide orders. Fees could be imposed on order cancellations. If small enough, they might reduce the huge number of order cancellations without causing significant problems for traders posting bona-fide orders. Markets could also require that orders not be canceled until one second (or some other time span) has passed after they are placed.
6. Many rogue traders need to expend significant amounts of time to cover their activities. Working late into the night provides them cover, enabling them to better obscure their activities, away from direct observation of their supervisors and staff monitoring their trading activity. Significant blocks of time away from the office help to prevent trader from consistently covering their activities, enable certain trading and reporting patterns to be broken, and provide monitor time to examine trading account histories in the absence of trading activity.
7. Front running is defined as the scenario in which a broker uses his knowledge of a client's pending order to buy (sell) a particular security to buy (sell) on his own account so as to benefit from the market reaction to his client's order. If HFT firms are not responding to their own client orders, they are not front running, although their orders may well have the same effect.
8. First, payment for order flow (maker-taker pricing or rebate programs) provides for trade execution, allowing for "immediacy" consistent with the Demsetz argument. Make-or-take pricing systems have reduced spreads and improved liquidity, even though the liquidity rebates normally are not reflected in bid and offer quotations. Rebate programs have the potential to help customers get price improvement for their orders by creating competition among the exchanges, market makers, and liquidity providers with whom they trade, potentially resulting in superior execution quality. In addition, the practice has virtually eliminated commissions and fees for many investors. This practice has continued to grow more widespread during the past 20 years.

Glossary

12(b)-1 fee Fee assessed shareholders by mutual funds for certain promotional expenses.

Abnormal return The part of a security's return that was not expected; that is, it appears to be attributable to something other than time value of money or risk. Also called excess return or residual.

Accredited investor An institutional or individual investor that is allowed by the SEC to invest in unregistered securities. Normally meets an income test (annual income exceeding $200,000, or $300,000 joint), is a "knowledgeable employee" of certain investment funds or holds valid FINRA licenses, or maintains a net worth level exceeding $1 million excluding the value of a primary residence.

Actual Size Rule SEC Rule reducing the minimum displayed quote size from 1000 to 100 shares.

ADR See American depository receipt.

Adverse selection Precontractual opportunism where one party to a contract (e.g., purchaser of insurance) uses his private information (e.g., information with respect to risk-taking) to the other counterparty's disadvantage.

Agency issue A security issue by an agency of the U.S. federal government. Includes securities issued by GNMA (Ginnie Mae), FNMA (Fannie Mae), and SLMA (Sallie Mae).

Agency trader Trader who acts as a commission broker on behalf of clients.

Algo trading See Algorithmic trading.

Algorithmic trading (1) Automated trading with the use of live market data and rule-driven computer programs for automatically submitting and allocating trade orders among markets and brokers as well as over time so as to minimize the price impact of large trades. Also called automated trading, black box trading, and robotrading. (2) Alpha model automated trading that is used to make trade decisions to generate profits or control risk.

Alpha (1) See Jensen's alpha. (2) Model, program, or strategy intended to generate a profit.

Alpha model (1) Gauges the performance of a trading strategy relative to a benchmark on a risk adjusted basis. (2) Model that seeks to outperform a benchmark on a risk-adjusted basis.

Alternative trading system (ATS) A securities trading venue that is not registered with the SEC or other national authority as an exchange.

American depository receipts (ADRs) Shares issued by banks evidencing ownership of shares of foreign company stock that are held by its sponsor through a trust.

American option Can be exercised any time prior to expiry.

Arbitrage (1) The simultaneous purchase and sale of the same asset. (2) The near-simultaneous purchase and sale of assets generating nearly identical cash flow structures.

Arbitrage opportunity Arbitrage transaction producing a profit, by purchasing at a price that is less than the selling price.

Arbitrageur Trader who seeks to profit from arbitrage opportunities.

Arrival price (bid−ask midpoint or BAM) The midpoint of the bid−offer spread at the time the order is received.

Ask A solicitation to sell (also called an offer).

ATS See Alternative trading system.

Auction A competitive market process involving multiple buyers, multiple sellers, or both. An auction is the process of trading a security through bidding, then selling it to the winning bidder.

Back office Departments or functions that provide indirect support to front office activities. Back office functions handle clearing and settlement as well as delivery and regulatory procedures and include personnel such as IT specialists, accountants, operations staff, human resources staff, office managers, customer service representatives, and regulatory compliance staff.

Backwardation The market condition in which a forward or futures contract price is below the underlying instrument's spot price, with the two converging at contract settlement.

BAM See Arrival price.

Banging the close Flooding a market sell orders its final minutes of trading before the close, thereby forcing prices down as buying interest diminished. A form of wash sale intended to manipulate prices.

Banker's acceptance Marketable security originated documenting a bank's acceptance of a responsibility to pay a client's loan or for assuming some other financial responsibility on behalf of a client.

Bargaining The negotiation process over transaction contract terms that occurs between a buyer and a seller for a transaction.

Basis point 0.0001 of a unit. Typically used in foreign-exchange and fixed-income market quotes.

Basis The potential for a price to move in the direction anticipated by an arbitrageur, at least in the short run.

Basis risk Risk that markets might move too slowly to profit from an apparent arbitrage, or that markets might move opposite to the arbitrageur's expectations, at least in the short run.

Batch processing A transfer system that allows for both transaction debits and credits to accounts to accumulate for each institution, then settle as a batch at a later time or date.

Bear One who anticipates a decline in prices.

Benchmark rate A market or computed rate defined in a loan or derivative contract, largely to reflect updated market interest rate conditions.

Beneficiary The entity that receives financial benefit or other consideration from a contract or arrangement.

Bernoulli trial A single random experiment with two possible outcomes (e.g., 0, 1) whose outcomes depend on a probability.

Best execution Requirement that a broker-dealer exercise reasonable diligence to maximize the economic benefit to the client of each transaction, taking into account best price, order size, speed of execution, clearing costs, and the cost, risk, and difficulty of executing an order in the various markets.

Best price Requirement based on the S.E.C.'s Order Protection Rule that client orders be executed at their best quoted prices (the NBBO).

Beta Coefficient that measures the risk of a security relative to the risk of some factor (usually the market).

Bid A solicitation to buy.

Bid-ask midpoint See Arrival price.

Binomial option pricing model Valuation model based on the assumption that the underlying stock follows a binomial return generating process.

Binomial process The asset value takes on only one of two potential constant values at any given time period.

Bitcoin A peer-to-peer (P2P) payment network and digital currency that does not rely on a central monetary authority to issue, monitor, facilitate, verify, and approve transactions. Instead, Bitcoin is underwritten by a peer-to-peer network.

Black-Scholes option pricing model A continuous time−space option pricing formula.

Blockchain An immutable electronic alternative to a traditional ownership ledger, based on digital time stamping of documents in chronological sequence used to authenticate authorship of property. This type of distributed ledger, comprising a chain of permanent, digitally recorded transaction data in packages called blocks, is used to identify and record ownership and transactions.

Blockholder An owner of a significant number of shares of a firm. Sometimes refers to an owner of 10,000 or more shares.

Block trade Transaction involving more than 10,000 shares or $200,000 of a security.

Bluffing The act of fooling other traders into making unwise trades by convincing them that the bluffer has superior material information.

Boiler room An enterprise typically operated in an inexpensive facility and that likely uses high pressure sales tactics and sometimes false or misleading information to solicit investors.

Bond Financial security that makes fixed payment(s) at specified interval(s).

Broker A security market participant who acts as an agent for investors, buying and selling on their behalf on a commission basis.

Brownian motion A Newtonian nondifferentiable stochastic continuous time−space process whose increments are independent over time.

Buy side traders Traders who buy liquidity services in the marketplace. Such traders often include individual investors, mutual funds, and pension funds.

Call (1) Security or contract granting its owner the right to purchase a given asset at a specified price on or before the expiration date of the contract. (2) In a call market, the process of engaging a trading session.

Call market Market where orders can be executed only at specific points in time, when orders are called (when buy orders are matched against sell orders or when orders are accumulated before being executed).

CDO See Collateralized debt obligation.

CDS See Credit default swap.

Central limit order book (CLOB) A system that matches anonymous limit orders.

CFTC See Commodities Futures Trading Commission.

Charting The process of using historical price and volume patterns from charts or graphs to devise investment trade strategies.

Chartist An investor or trader who uses patterns from charts or graphs of a financial instrument's historical prices and volume to devise trade strategies.

Clawback A compensation provision that allows for already vested or paid employee compensation to be reclaimed by an employer based on gross negligence or other malfeasance on the part of the employee.

Clearance Recording and comparison of records associated with a trade.

Clearing See Clearance.

Clearing firm Firm authorized by a clearing house to manage trade comparisons and other back office operations.

Clearing house Institution that clears transactions for an exchange or other market.

Closed-end investment company Investment company that issues a specified number of shares that can be traded in secondary markets such as the New York Stock Exchange.

Coefficient of correlation A measure of the strength and direction of the relationship between two sets of variables. Ranges between -1 and $+1$ and may be regarded as a "standardized" covariance (dividing covariance by the product of the standard deviations of the two variable sets).

Collar A portfolio consisting of a long position in a put and a short position in a call, collaring the underlying security selling price to lie between the exercise prices of the two options contracts.

Co-location The practice by high-frequency traders to rent space from or very near market centers, such as the major exchanges ECNs and other and alternative trading systems for the purpose of co-locating their servers next to the market center's data servers to reduce latency and access to information. Also called proximity hosting.

Collateralized debt obligation (CDO) An agreement or marketable instrument by which specified events determine the payouts associated with debt-backed assets.

Commodities Futures Trading Commission The U.S. federal agency operating under the Department of Agriculture that regulates futures and commodities trading in the United States, including the trading of FX contracts. Also called the CFTC.

Common stock Security held and potentially traded by the residual claimants or "owners" of the firm.

Common value auction Auction where all bidders place the same value on the item to be auctioned, and that value is known with certainty.

Confirmation Verification of a trade.

Consolidated Quotations System Provides traders with price quotes from the various exchanges and FINRA and calculates the NBBO.

Consolidated tape High-speed electronic system for reporting transaction prices and volumes for securities on all U.S. exchanges and markets.

Continuous market Market where orders can be executed whenever the market is open.

Convexity (1) The slope of the slope of a function. (2) The sensitivity of the duration of a bond to changes in the market rate of interest.

Corner The purchase of a sufficient level of a given security to obtain market or monopoly power over its price.

Counterparty Person or institution with whom one trades.

Counterparty risk The potential that a trade counterparty fails to fulfill their side of a transaction.

Coupon The interest rate on debt as a percentage of its face value.

Covariance A statistical measure of the co-movement between two sets of variables.

Covered call writing Writing or selling a call while owning the underlying asset (or appropriate combination of other assets) needed to generate a hedge.

Covered put writing Writing or selling a put while shorting the underlying asset (or appropriate combination of other assets) needed to generate a hedge.

Cox Ingersoll Ross process A mean-reverting stochastic process defined by the differential $dr_t = \lambda(\mu - r_t)dt + \sigma_r\sqrt{r_t}dZ_t$.

Credit default swap (CDS) A derivative instrument intended to transfer default risk by providing for payment by the seller to the buyer in the event of default by a particular debtor.

Cross Leaving a bid (offer) unexecuted in one market that is greater (less) than an offer (bid) on the same security in another market. Prohibited by Reg NMS.

Crossed market Market where one or more bids exceed one or more offer prices.

Crossing network Alternative trading system that matches buyers and sellers with respect to agreed-upon quantities, and does not publicly display quotations, thereby enabling participants some degree of anonymity. Trades are priced by reference to prices obtained from other markets rather than by auctions. Traders using crossing networks announce their order sizes, which are matched at prices obtained auction markets such as the NYSE.

Cryptocurrency Digital currency in which encryption (cryptography) techniques are used to regulate its creation and supply and to verify its transfer among users.

Cumulative average residual (CAR) The sum of abnormal returns or residuals over a period of time around an event.

Cumulative distribution function A function F of a random variable so that $F(x)$ equals the probability that the random variable is no larger than x.

Current yield Measure of the annual interest payments made by a bond relative to the initial investment required by the bond.

Custodian Holds assets on behalf of owners.

Dark liquidity pool Broadly defined, any willingness to buy or sell that is not publicly displayed. Often refers to portions of large trades broken up into small trades to be executed in different markets and over time so as to minimize the price impact (slippage) of the trade. Sometimes refers to crossing networks.

Dealer Securities market participant who trades directly with clients and brokers, seeking trade counterparties for clients, facilitating their trading process, posting and maintaining quotes, and buying and selling for their own portfolios on a profit basis.

Default waterfall The sequence of resources available to a central clearing party or clearinghouse to protect the market and its participants from broker and clearing firm defaults.

Delta Sensitivity of an option's price to changes in the price of the underlying security.

Delta neutral Portfolio of options and underlying securities whose weighted average Black-Scholes delta is zero, and whose value is constant with respect to small changes in the value of the underlying asset.

Depository Trust & Clearing Corporation (DTCC) A holding company for U.S. clearing and settlement companies, including the Depository Trust Corporation, the Fixed Income Clearing Corporation, and the National Securities Clearing Corporation.

Depository Trust Company (DTC) A unit of the Depository Trust & Clearing Corporation (DTCC) that provides book-entry settlement and custody services for almost all publicly traded equity, debt (corporate and municipal), asset-backed, and money market securities traded in the United States.

Depth A market's ability to process and execute a large order without substantially impacting its price.

Derivative security Security whose payoff function is derived from the value of some other security, rate, or index.

Designated market maker (DMM) Exchange official and market participant charged with the responsibility to maintain a fair and orderly market in their stocks, quote at the NBBO a specified percentage of the time, and facilitate price discovery throughout the day as well as at the open, close, and in periods of significant imbalances and high volatility. The DMM provides price improvement and matches incoming orders based on a preprogrammed capital commitment schedule, which has been added to the NYSE Display Book and minimizes order latency. DMMs compete as market participants, trading on their own accounts.

Direct access Trading systems that provide for direct access to markets bypassing brokers.

Discount rate A rate used to discount (usually reduce) future cash flows to express their values relative to current cash flows.

Discrete process A process whose variable can be assigned only a countable number of values.

Distributed ledger Accounting records maintained on multiple distinct computing systems.

Diversification Holding multiple assets whose returns are not perfectly correlated.

DKs (don't knows) Trade reports filed with clearing firms or clearing houses with discrepancies resulting from recording errors, misunderstanding, and fraud.

DMM See Designated market maker.

Don't knows See DKs.

Double auction Used for the trading of most publicly traded securities in secondary (resale) markets in which buyers submit bids (maximum purchase prices) and sellers submit offers (minimum selling prices) that are ranked from best to worst. Also called bilateral auctions.

Dow Jones Industrial Average (DJIA) A price-weighted average of 30 actively traded blue chip stocks.

Drift The predictable change component of a stochastic process.

DTC See Depository Trust Company.

DTCC See Depository Trust & Clearing Corporation.

Duration Measures the proportional sensitivity of a bond value or price to changes in the market rate of interest.

Dutch auction Auction that begins with the auctioneer calling out a high offer price, which is reduced until some participant submits the first, highest, and winning bid. Also called descending bid auction.

ECN See Electronic communication network.

Effective spread The quote midpoint at the order arrival relative to the actual execution price. Typically measured as the absolute value of the execution price minus the NBBO midpoint multiplied by two.

Efficient market A market where security prices fully reflect all available information.

Electric communication network Virtual meeting place and screen-based system for trading securities.

Eligible counterparties (ECP) A European (MiFID) category of sophisticated investors composed of investment firms, credit institutions, insurance companies, other financial institutions, central banks, and national governments deemed less needy with respect to regulatory protections.

English auction Auction where participants bid openly against one another, with each successive bid higher than the previous one. Also called ascending bid auction.

Equity Security that represents ownership in a business or corporation. Normally called stock.

ETF See Exchange traded fund.

EUREX A German-Swiss derivatives exchange created by the merger of the German and Swiss exchanges; part of Deutsche Börse.

Eurobond Generally underwritten bearer bond denominated in a currency other than that of the country where the loan is extended.

Eurocredit Bank loans denominated in currencies other than those of the country where the loan is extended.

Eurodollar Freely convertible dollar-denominated time deposit or debt instrument issued outside the United States and outside of the jurisdiction of the U.S. Federal Reserve.

EURONEXT A pan-European securities and derivatives exchange listing Dutch, French, Portuguese, and Belgian securities and derivatives plus derivative products traded on LIFFE.

European option Can be exercised only when it expires.

Exchange A physical or electronic marketplace for trading securities.

Exchange-traded fund (ETF) Closed-end investment company whose shares are traded on an exchange.

Execution The process of selling or buying a security.

Execution speed The time it takes for an order to be executed.

Exercise price Price at which an option can be exercised; that is, the price that the call owner has the right to pay for the underlying asset or put owner has the right to sell the underlying asset for. Also called striking price.

Expectations theory See Pure expectations theory.

Face value The principal or par value of debt.

Federal funds rate The rate at which the excess reserves of one bank can be loaned to other banks for satisfaction of borrower reserve requirements.

Fiduciary A person or institution that maintains a legal or ethical relationship of trust with another, typically managing or safekeeping assets and acting in the best interests of the client.

Fiduciary standard Holds that an agent or advisor (or Registered Investment Advisor) must put her client's interests above her own, following the very best course of action for the client, regardless of how it affects the advisor personally. That is, the agent or advisor's primary loyalty is to the client.

Fill or kill order A securities market order that must be filled in its entirety immediately, or it is canceled.

Filtration History, or indexed set of subobjects in a stochastic process.

Financial Institutions Regulatory Authority (FINRA) The largest independent regulator for all securities firms and exchanges conducting business in the United States. Formed by the merger of the NASD, and member regulation, enforcement, and arbitration operations of the New York Stock Exchange.

Financialization An economy-wide phenomenon in which profits accumulate primarily through financial channels rather than through trade and commodity and service production.

FINRA See Financial Institutions Regulatory Authority.

First-price sealed-bid auction All bidders simultaneously submit sealed bids so that no bidder knows any of the other bids. Does not allow for price discovery until the auction concludes. The winner submits the highest bid and pays the bid price.

Fisher effect Defines the real interest rate as the ratio of the nominal rate to the inflation rate.

Fishing Practice intended to obtain secret information concerning order sizes in dark pools. Occurs when a trader sends a series of small orders to a dark pool to detect whether there is a large order waiting in that pool.

Flash order Quote that is immediately executed or canceled (within 500 milliseconds of the original quote), need not be publicly displayed, and may be displayed to select clients. Flash quotes are the basis for flash trading.

Flash trading Trading motivated by private or select market participant receipt of quotes (flash orders) in advance of the public quote stream (also called prerouting displays). This means that the recipients of flash orders receive access to quotes before the public and can engage in flash trading on the basis of these quotes.

FHLMA Federal Home Loan Mortgage Association. Founded in 1969 by the federal government of the United States to facilitate mortgage lending to prospective homeowners. It provides capital for the mortgage industry by selling securities backed by other mortgages issued by banks.

FNMA Federal National Mortgage Association (Fannie Mae or FNMA). Founded in 1938 by the federal government of the United States to facilitate mortgage lending to prospective homeowners. It provides capital for the mortgage industry by selling securities backed by other mortgages issued by banks.

Foreign exchange Pertaining to trading of currencies.

Foreign exchange option Option on currency.

Forward contract Instrument that obliges participants to either purchase or sell a given asset at a specified price on the future settlement date of that contract.

Forward rate Locked-in interest rate on loans originating after time zero or in the future.

Fragmentation The condition in which multiple market venues exist; the opposite of a concentrated market.

Front running Using the information in another participant's order to set a bid or quote in a subsequent order with priority. An example of front-running might be where a broker with information about a client's order uses this information to set a bid or offer in her own order in setting a bid or offer so as to exploit the client's order.

Futures contract Instrument that obliges participants to either purchase or sell a given asset at a specified price on the future settlement date of that contract. Normally traded on an exchange, with standardized contractual terms and providing for margin and marking to the market.

Gamma Sensitivity of an option's delta to changes in the price of the underlying security.

Geometric Brownian motion A stochastic process whose logarithm is a sum of Brownian motion and drift.

GNMA Government National Mortgage Association (Ginnie Mae or GNMA). Founded by the federal government of the United States to facilitate mortgage lending to prospective homeowners. It guarantees mortgage-backed securities issued by banks and other institutions.

Greeks Sensitivities of options and portfolios of options to Black-Scholes pricing model inputs.

Haircut (1) The difference between the market value of a security and its value as collateral. (2) The dealer's spread on a tradable instrument. (3) The reduction or potential reduction in value received by or to be received by a creditor in the event of a default.

Hedge To take a position to reduce risk.

Hedge fund A private fund that allows investors to pool their investment assets. To avoid SEC registration and certain regulations, hedge funds usually only accept funds from small numbers (often less than 100) of accredited investors, typically high-net-worth individuals and institutions.

Hedge ratio Defines the number of units of one security required to offset the position in one unit of another security in order to form or maintain a riskless portfolio.

Hedging probability See Risk-neutral probability.

HFT See High frequency trading.

High frequency trading The practice of rapid executions of multiple transactions for securities followed by extremely short holding periods, perhaps as short as fractions of seconds.

Hot potato trading Repeated passing of inventory imbalances among dealers.

Hypothecation The pledge of assets as collateral for debt or other obligation, typically in scenarios involving mortgages, margin, short sales, and repo agreements.

Iceberg order Partially revealed order whose full size is not revealed. Also called hidden-size order.

Immunization Strategies concerned with matching the present values of asset portfolios with the present values of cash flows associated with future liabilities. More specifically, immunization strategies are primarily concerned with matching asset durations with liability durations.

Implementation risk Risk taken by arbitrageur that one or more arbitrage transactions might fail to execute or be executed at prices that differ from what the arbitrageur anticipated.

Implementation shortfall The performance difference between hypothetical profits realized by a paper portfolio replicating an actual portfolio and the profits realized by the actual portfolio.

Implied volatility Variance backed out of an option pricing model using market prices.

Improve (a market) To place a bid higher than the best bid or an offer less than the best offer.

Index A portfolio of stocks or other instruments intended to reflect performance of a particular market or sector.

Infinitesimal Value approaching zero.

Information asymmetry Scenario where agents have information that differs in quality.

Inside information Material nonpublic information concerning the prospects of a security.

Interest rate parity States that anticipated currency exchange rate shifts will be proportional to countries' relative interest rates.

International Fisher effect By combining purchase power parity, interest rate parity, and the Fisher effect, the international Fisher effect demonstrates that different economies will have identical real interest rates.

Internalize Fill a broker-dealer customer's order by the broker-dealer itself without assistance of an exchange or ATS.

Investment company Institution that accepts funds from investors for the purpose of investing on their behalf.

Intermarket trading system Displays quotes in different markets and links markets for trade executions to facilitate investors' access to the best quotes.

Investment company An institution that accepts funds from investors for the purpose of investing on their behalf.

Jensen's alpha Excess return on a portfolio over the CAPM-implied risk-adjusted return.

Jitney game A series of wash sales at ever-increasing prices.

Kyle's lambda Measures the illiquidity of a security.

Latency The amount of time that lapses from when a quote or an order is placed by a trader and when that order is actually visible to the market.

Late trading An illegal fund trading strategy purchase or redeem shares after daily net asset values (NAVs) are computed.

Law of one price States that securities or portfolios of securities offering the same cash flow characteristics or baskets offering the same commodities must sell for the same price.

Layering Submission of a large number of quotes with the intention of canceling them after creating the false appearance of a market's depth or direction.

Leg One of the single-contract option components of an option spread.

Leveraged buyout The purchase of a firm by private group of investors with substantial debt financing. It is usually a going private transaction; that is, a publicly traded company is transformed into a privately owned company in the transaction.

Level I quotes Displays the best bid and offer prices (inside quote or BBO, best-bid-offer) and, in some cases, quote sizes.

Level II quotes Display the best bid and offer prices, quote sizes along with other quotes in descending order for the best bids and ascending for the best offers with market maker symbols for each.

Level III quotes Offered to Nasdaq members, provide the same information as Level II and enable traders to have direct access to enter and revise quotes.

LIBOR (London Interbank Offered Rate) A quoted benchmark interest rate widely used when pricing derivative securities or setting loan rates.

Limit-on-close (LOC) An order to execute at the daily closing auction only if the closing price is at least as good as the limit price set by the client.

Limit order Sets an upper price for a purchase or a lower price limit for a sell, preventing the broker from paying more or accepting less for the security. The limit order restricts the price of the order.

Limit Order Display Rule Requires that sizes and prices of customer limit orders be displayed in the NBBO when they are better than market maker quotes.

Limit up-limit down (LULD) An SEC rule prohibiting trading in exchange-listed securities at prices outside upper and lower bands set at given percentages above and below the average price of the security during the immediately preceding 5 minutes.

Liquidity An asset's ability to be easily purchased or sold without causing significant change in the price of the asset.

Liquidity rebate See Payments for order flow.

Load fund A mutual fund that charges a sales fee.

Lock Leaving a bid (offer) unexecuted in one market that equals an offer (bid) on the same security in another market. Prohibited by Reg NMS.

Locked market A scenario in which the bid in one market equals the offer in another.

Long (1) To have purchased or owned (2) position that obligates or enables the investor or participant to purchase a given or underlying asset on the expiration or settlement date of a derivative contract.

Maintenance margin Adjustment to the minimum equity an investor must hold in margin after a transaction is executed.

Maintenance requirement The equity level that must be maintained in a margin account, as required by NYSE, FINRA, individual brokerage firms, and other regulatory authorities.

Make-or-take pricing Refers to exchanges allowing patient traders to post standing limit orders that await execution until some other trader takes the other side, providing liquidity rebates to the patient trader making the market.

Maker-taker pricing See make-or-take pricing.

Management buyout The purchase of a public firm by the management of that firm. It is usually a going private transaction.

Margin Funds posted up front by traders to ensure they fulfill their trading obligations.

Market architecture The set of rules governing the trading process.

Market impact The effect that a given transaction has on the market price of a security.

Market-maker A trader or firm that maintains an expressed willingness on an ongoing basis to trade a given security at specified quotes.

Market microstructure Area of financial economics concerned with trading and market structure, market fairness, success and failure, and how the design of the market affects price formation and discovery. Market microstructure is concerned with costs of providing transaction services along with the impact of transactions costs on the short-run behavior of securities prices.

Market on close (1) An order placed to execute at the daily closing auction irrespective of price. (2) The last price obtained by a trader at the end of the day relative to the last price reported by the exchange. Used for trader evaluation purposes.

Market order An order to buy at the best offer or sell at the best bid.

Market timing A fund trading strategy intended to exploit stale security prices and net asset value (NAV) deviations from fundamental values.

Markov process (random walk) A stochastic process whose increments or changes are independent over time.

Mark to market Accounting for the fair value of an asset or liability based on its current market price, or, if unavailable, based on either similar instruments or an appropriate valuation model.

Martingale process A stochastic process whose increments have expected value 0.

Matching engine Software algorithm designed to enable exchanges to match buy and sell orders for execution.

Matter under inquiry (MUI) An informal or preliminary investigation by the SEC to determine whether a formal investigation is warranted.

Maturity The date at which payments cease on a debt security.

Method of bisection An iterative method of bisecting intervals to find the roots of an equation.

Microsecond One millionth of a second.

Millisecond One thousandth of a second.

MOC See Market on close.

Model risk Failure to fully understand the implications of a trading model.

Multilateral Trading Facility European variation on the alternative trading system.

Multiplicative movement The proportion by which an asset changes over a given time period.

Mutual fund (open-end investment company) An institution registered with the SEC under the Investment Act of 1940 that pools investors' funds into a single portfolio.

Naked access Where algorithmic trade orders do not pass through pre−trade-risk controls, potentially leading to runaway executions.

Naked call writing Writing or selling a call without owning the underlying asset (or appropriate combination of other assets needed to generate a hedge).

Naked put writing Writing or selling a put without shorting the underlying asset (or appropriate combination of other assets) needed to generate a hedge.

Naked short Short selling a security without borrowing it.

Nanosecond One billionth of a second.

NASD See National Association of Security Dealers.

National Association of Security Dealers (NASD) Formerly the primary independent self-regulatory organization for the securities industry responsible regulation of the Nasdaq stock market and over-the-counter markets and for administrating competency exams for investment professionals. Merged with the member regulation, enforcement, and arbitration operations of the New York Stock Exchange to form FINRA.

National best bid and offer (NBBO) The highest bid and lowest offer prices available on a security at a given point in time.

Nationally Recognized Statistical Rating Organization (NRSRO) A credit rating agency whose ratings the U.S. Securities and Exchange Commission has designated as authoritative for certain regulatory purposes.

NBBO See National best bid and offer.

Negotiable certificate of deposit Tradable depository institution certificate of deposit with a denomination exceeding $100,000.

Net asset value (NAV) Market value of portfolio assets minus its liabilities at a particular point in time divided by the number of shares in the portfolio outstanding at that time.

Netting The simplification process used by clearing institutions of adding all of a given firm's purchases of each security, adding the sales of each security, deducting sells from buys to determine the net change in holdings of that security for the firm, and computing the net cash flows associated with all transactions.

Newsreader algorithms Electronic-based rules based on text mining and statistical techniques to analyze news sources, blogs, tweets, and other data to obtain relevant trading information and infer its impact on security prices.

Noise trader Trades on the basis of what he falsely believes to be special information or misinterprets useful information concerning the future price or payoffs of a risky asset.

No-load fund Mutual fund that accepts investments directly from investors without a sales charge.

Normal return The return on a stock or other asset sufficient to compensate investors for time value of money and risk. Often assumed to be a risk-adjusted rate of return or return of a comparable asset.

Not held order A securities market order in which the broker is not obliged to execute while he is attempting to obtain a better price for his client.

Novation The process whereby a clearing house assumes the settlement obligations of both counterparties to a transaction, in effect becoming the counterparty to both sides of the transaction.

NRSRO See Nationally Recognized Statistical Rating Organization.

OARS See Opening Automated Report Service.

Offer A solicitation to sell (also called an ask).

Opaque markets Markets that lack transparency.

Open-end investment company See Mutual fund.

Opening Automated Report Service (OARS) NYSE system that allows orders that have accumulated overnight or while the exchange is closed to be matched so that the designated market maker attempts to establish a market clearing price.

Option Security that grants its owner the right to buy (call) or sell (put) an asset (underlying asset) at a specific price (exercise or striking price) on or before the expiration date of the option contract.

Order-driven markets Markets where traders can trade without the intermediation of dealers.

Order Handling Rules (1) Generally, rules that concern a market's display, routing and execution of orders. (2) More specifically, the SEC's Limit Order Display Rule, the Quote Rule, and the Actual Size Rule.

Order precedence Rules that determine which traders can place bid and offer quotations with greatest priority for execution and which can accept the quotations of other traders.

Order Protection Rule SEC Rule 611, also known as the "trade-through" rule, requiring that exchanges, ATSs and OTC market makers maintain and enforce policies to ensure trade executions at best prices.

Ornstein-Uhlenbeck process A mean reverting stochastic process defined by the differential $dr_t = \lambda(\mu - r_t)dt + \sigma dZ_t$.

Out trades Trade reports filed with futures clearing firms or clearing houses with discrepancies resulting from recording errors, misunderstanding, and fraud.

Over-the-counter (OTC) markets Securities markets other than those provided by exchanges.

Pairs trading An arbitrage strategy that involves taking offsetting positions in two different stocks (perhaps options or index contracts) with correlated returns, one long and one short, such that gains in one position are expected to more than offset losses in the other position.

Parasitic trading Broker trading on the basis of knowledge of her client's trading activity.

Pattern day trader SEC definition for a trader who executes four or more round trip transactions in the same account within five consecutive business days.

Payment for order flow Compensation received by brokers and other institutions from exchanges and alternative trading systems as payment for directing their order flow to those markets.

Penny jumping A type of parasitic trading where the broker places a buy (sell) order one uptick above (downtick below) below the client's buy (sell) limit order, expecting to benefit either from the market's reaction to the client order or to limit his losses by transacting with the client.

Pick off To force a bad trade by a counterparty, often by hitting the counterparty stale or erroneous quote.

Ping To place a small order to detect the level of liquidity and possibility of large orders. If small orders are quickly executed, there might be significant interest in the security.

Ping destination A pool, typically operated by a hedge fund or electronic market maker, that accepts only immediate or cancel (fill or kill) orders.

Pip Basis point, or 0.0001 of a unit. Typically used in foreign exchange markets quotes.

Ponzi scheme A type of fraud in which belief in the success of a nonexistent enterprise is enhanced by payments to early investors from money invested by later investors.

Portfolio A collection of investment holdings.

Position An asset held or portfolio obligation incurred, normally as a result of a transaction.

Precedence Which traders can execute bid and offer quotations and which can accept and execute the quotations of other traders. In most cases, the primary precedence is price priority; the participant with the highest bid or lowest ask has priority on an execution.

Predatory algo trading Strategies designed to exploit other institutional algo orders.

Prediction market A trading venue created to predict an outcome based on prices of specially created and traded securities that pay contingent on that outcome.

Price discovery The process of determining the worth or price of an asset in the marketplace through the interactions of buyers and sellers.

Price improvement Providing a better execution price than the quoted NBBO on market and limit orders; that is, providing a higher bid or lower offer price than the posted NBBO.

Price priority Rule or rules dictating that the highest bids or lowest offers will be executed against in the presence of competing quotes.

Priority Rule or rules dictating which quotations are to be executed in the presence of competing quotes.

Private equity Refers to asset managers that make equity investments in companies that are not publicly traded.

Probability distribution A rule or an equation that assigns a probability to every possible value or range of values of a random variable.

Program trading Defined by the NYSE as computer-initiated trades involving 15 or more stocks with value totaling more than $1 million.

Prop shop See Trading arcade.

Prop trader See Proprietary trader.

Proprietary trader Trader *who* seeks profits by trading on his own account.

Proprietary trading firm See Trading arcade.

Protected quotes By Regulation NMS, the best publicly displayed bid and offer on the exchanges or OTC market, but only to the extent those quotations can be automatically accessed. Cannot be bypassed for execution in favor of a worse quote.

Proximity hosting See Co-location.

Public Company Accounting Oversight Board (PCAOB) Nonprofit corporation created by Sarbanes-Oxley to oversee public auditing firm practices.

Pump and dump Market manipulator touts a security with false and misleading statements, causing the price to rise before selling their own holdings at the inflated prices.

Purchase power parity States that commodities selling in different countries must sell for the same price after adjusting for exchange rates.

Pure discount note A debt security paying no interest; it only pays its face value or principal.

Pure expectations theory Term structure theory that states that long-term spot rates can be explained as a geometric mean of short term spot and forward rates.

Put Security or contract granting its owner the right to sell a given asset at a specified price on or before the expiration date of the contract.

Put-call parity An equation that expresses the pricing relationship between a put and a call.

Qualified institutional buyer (QIB) Institution with an investment portfolio exceeding $100 million.

Quote A solicitation to buy (bid) or solicitation sell (offer or ask).

Quote-driven markets Markets where dealers post quotes and participate on at least one side of every trade.

Quote matching Occurs when a small trader places an order one uptick (downtick) from that of a large trader so as to profit from the large trader's transaction upward (downward) price pressure, or to use the large trader as a trade counterparty should prices decrease (increase).

Quote Rule Formally referred to as SEC Rule 602 of Regulation NMS, requires that all market centers publicly disseminate their best bids and offers through the securities information processors (such as the Consolidated Tape, Consolidated Quotations System, and the Intermarket Trading System).

Quote stuffing The placement of large numbers of rapid-fire stock orders, with most or all of which being canceled almost immediately, frequently for the purpose of clogging trading HFT and other algorithms and data computations.

Random walk A process whose future behavior, given by the sum of independent random variables, is independent of its past.

Real estate investment trust (REIT) A fund that invests in real estate and/or real estate mortgages, providing investors with opportunities to diversify into real estate with relatively small investment sums.

Realized spread The difference between an execution price and the quote mid-point at some point in time after the transaction (typically a few minutes, often depending on market transparency).

Recombining Multiple paths lead to the same event or outcome.

Reference rate See Benchmark rate.

Regulation T A Federal Reserve Board regulation covering extension of credit to clients by securities brokers, establishing initial margin requirements and defining eligible securities.

Rehypothecation The repledging of assets accepted in a hypothecation agreement from a borrower as collateral for an additional debt by the original lender.

Repo See Repurchase agreement.

Repurchase agreement (repo) Marketable security issued by a financial institution acknowledging the sale of assets and a subsequent agreement to repurchase at a higher price in the near term. This agreement is essentially similar to a collateralized short-term loan.

Residual claim The right to take possession of assets unclaimed by others.

Resiliency The ability of a market to be unaffected by or bounce back after a large order or transaction.

Return Profit relative to the initial investment amount.

Revenue equivalence theorem States that under specific restrictions, the auction type will not affect auction outcomes.

Rho Sensitivity of an option's price to changes in the riskless return rate.

Risk (1) Uncertainty. (2) The likelihood of loss, failure, or another undesirable outcome.

Risk free rate The return or interest rate on an asset with no risk.

Risk-neutral probability The probability in a complete market that guarantees arbitrage-free pricing and in which the discounted value of any security is a martingale.

Risk premium Return offered or demanded as compensation for accepting uncertainty.

Robotrading See Algorithmic trading.

Rogue trader A trader who acts independently of others and the authority of her employer, typically in a reckless manner, frequently to the detriment of her clients and employer.

Routing The process of executing the trade, involving broker selection, deciding which market(s) will execute the trade(s), and transmitting the trade(s) to the market(s).

Rule 10b-5 (SEC) Prohibits any act or omission resulting in fraud or deceit in connection with the purchase or sale of any security.

Rule 144A market Market specifically created for "qualified institutional buyers" to trade unregistered securities held by less than 500 shareholders.

SDBK See Super Display Book System.

Seat A contract or arrangement granting trading privileges on an exchange.

SEC See Securities and Exchange Commission.

Secondary market A market in which securities are bought and sold among investors.

Second-price sealed-bid auction All bidders simultaneously submit sealed bids so that no bidder knows any of the other bids. Does not allow for price discovery until the auction concludes. The winner submits the highest bid and pays the second-highest bid price (the highest losing bid). Also known as a Vickrey auction.

Secured Overnight Financing Rate (SOFR) A benchmark or reference rate based on overnight repurchase agreement (repo) transactions. The volume-weighted median interest rate calculated by the Federal Reserve Bank of New York from repo transactions that reflects circumstances where credit, liquidity, and other risks are minimal.

Securities and Exchange Commission (SEC) The U.S. Federal regulatory agency established in 1934 to administer Federal securities laws.

Securities Industry Automation Corporation (SIAC) A subsidiary of the NYSE that provides a variety of technical services to exchanges, members, and other financial institutions to support trading, order processing, and reporting of market data.

Securities Information Processor (SIP) (1) System overseen by the CTA and maintained by the NYSE to consolidate and disseminate quotes and execution prices, and to calculate the NBBO from trade venues into a single data feed available to all broker/dealers. (2) Analogous systems maintained by other markets and trading firms.

Security Tradable claim on the assets of an institution or individual.

Self-regulatory organization (SRO) A registered exchange or registered securities or futures association that enforces financial and sales practice requirements as per Federal regulations.

Sell side traders Traders who sell side liquidity services, buying or selling when others need to execute transactions. Such traders normally include day traders, market makers, and brokers.

Semistrong form efficiency Scenario where prices reflect all publicly available information.

Separate Trading of Registered Interest and Principle Securities (STRIPS) The U.S. Treasury's program for debt instruments on which coupon and principal payments can be detached and marketed separately (see Strip).

Settlement Physical or electronic actual delivery of the security and payment involved in a trade.

Settlement Date (1) The date at which a financial instrument is physically delivered or delivered for book entry and at which payment for said instrument is made. (2) The date specified in a forward or futures contract at which a commodity or financial instrument will be delivered for payment.

Settlement Price (1) The price defined by a forward or futures contract at which a commodity or financial instrument will be exchanged at some future date, known as a settlement date. (2) In derivatives markets, the price used for determining profit or loss for the day and for calculating margin requirements. That is, the settlement price might be the final forward or futures trade price for that day.

Sharpe ratio Index of risk-adjusted portfolio performance, calculated as the portfolio's excess return to the risk-less rate relative to its standard deviation.

Short (1) To have sold or short sold (2) position that obligates or enables the investor or contract participant to sell a given or underlying asset on the expiration or settlement date of a derivative contract.

Short sell To sell a security without actually owning it. Normally requires its purchase at a later date and may involve borrowing the security from another investor first, returning it when it is repurchased at a later date.

SIMFU See Systemically important financial market utility.

Simple time slicing Order is split up and sent to markets at regular time intervals.

SIP See Securities Information Processor

Slippage Unwanted price impact as a trader's buy pressure forces the price up or his sell pressure forces the price down.

SOFR See Secured overnight financing rate.

SPAN See Standard portfolio analysis of risk.

SPDR Highly liquid ETF based on the value of the S&P 500 index.

Speculate To take a position of risk based on a forecast of the direction of a security price change.

Speculator Trader who speculates.

Spider See SPDR

Sponsored access Where brokers loan their market participation identification numbers (MPIDs) to high-frequency trading traders, enabling them to gain faster, direct access to markets by bypassing the broker's trading systems.

Spoofing Placing a quote that is intended to be canceled prior to execution.

Spot rates Interest rates on loans originating at time zero or now.

Spot transaction Occurs at the time of the agreement to make the exchange.

Spread The difference between the best offer and bid prices.

Squeeze An effort to force security purchases at low prices, usually used in the context of short sellers who must repurchase short-sold securities.

SRO See Self-regulatory organization.

Stale Not reflecting the most recent information, e.g., stale price or stale quote.

Standard Deviation The square root of the variance. A measure of risk, uncertainty, dispersion or volatility.

Standard Portfolio Analysis of Risk (SPAN) A risk and margin calculation system used and adapted by many clearinghouses for use with futures portfolios to analyze a grid or array composed of possible changes in market conditions within boundaries of the risk parameters set by the clearinghouse. The array is structured to calculate the worst possible loss scenario for the portfolio. This analysis is then used to set the margin requirement for the portfolio.

Stat-arb See Statistical arbitrage.

Statistical arbitrage (stat-arb) Strategies that seek to exploit mispricing of one or more assets based on their expected values. By using statistical methods, the arbitrageur takes positions in large numbers of trades, expecting to gain from expected mispricing through the law of large numbers.

Stealth trading Trading with an effort to remain hidden or to hide intentions, often by spreading trades to different markets and at different times.

Stochastic Random.

Stochastic process A sequence of random variables x_t indexed by time t. In other words, a stochastic process is a random series of values xt sequenced over time.

Stock (1) Equity, or residual ownership in a firm. (2) Security evidencing a proportional claim on a firm's residual ownership.

Stop-close-only order A stop order that can be executed, if possible, only at the close of the market.

Stop order Instructs the broker to place the buy order once the price has risen above a given level or place the sell order once the price of the security has fallen beneath a given level. The stop order triggers an order execution.

Stop limit sell order Authorizes the broker to initiate the sell order once its price drops to the stop trigger, but only if the limit price can be realized for the sale.

Straddle A combination of put and call options with identical exercise prices on the same underlying asset.

Strike price See striking price.

Striking price Price at which an option can be exercised; that is, the price that the call owner has the right to pay for the underlying asset or put owner has the right to sell the underlying asset for. Also called exercise price.

Strong form efficiency Scenario where prices fully reflect all information, public or private.

Stub quotes Quotes submitted with no intention of execution, often for the purpose of quote stuffing.

Strip Single payment fixed income instrument issued through the U.S. Treasury's Separate Trading of Registered Interest and Principle Securities (STRIPS) program.

Submartingale process A stochastic process whose increments have expected value greater than 0.

Suitability standard Holds that agents or advisors must provide clients with advice suitable to the client's needs, but not necessarily the best advice, as long as it is suitable. The advisor must believe that the advice benefits the client. However, the advisor's first loyalty might still be to her employer or to herself.

Supermartingale process A stochastic process whose increments have expected value less than 0.

Sunshine trading Occurs when traders announce their intentions, perhaps in an effort to increase the competition to act as counterparties on the same transactions, but when additional competition on the same side of the transaction is either unexpected or won't hurt the trader.

Super Display Book System (SBDK) The New York Stock Exchange ECN, order routing, and processing system.

SuperMontage Nasdaq's automated exchange.

Swap contract Instrument that provides for the exchange of cash flows associated with one asset, rate, or index for the cash flows associated with another asset, rate, or index.

Swap execution facility (SEF) A trading system or platform that enables traders to execute swap transactions by accepting bids and offers. A SEF can be similar to a formal exchange, although the term refers to a distributed group of approved trading systems for swaps.

Synthetic probability See Risk-neutral probability.

Systemic risk (1) Uncertainty or potential losses that pose a threat to the financial system. (2) Systematic risk.

Systematic risk Risk characterized by or substantially related to risk of the market as a whole or some identifiable grouped component thereof.

Systemically important financial institution (SIFI) A bank or other financial institution whose failure has the potential to trigger a financial crisis.

Systemically important financial market utility (SIFMU) A nonbank financial institution providing transfer, clearing, and settlement services whose failure has the potential to trigger a financial crisis.

Synthetic securitization The process of separating contingencies such as credit default from instruments of original issue (e.g., mortgages) then repackaging into instruments (e.g., credit derivatives) or portfolios to be marketed separately.

TAAPS See Treasury Automated Auction Processing System.

Tailgating A form of parasitic trading where the broker places an order immediately after the client's order.

Tâtonnement A preliminary iterative auction process to determine supply and demand levels at various prices.

Technical analysis Concerned with finding situations where historical price sequences can be used to forecast price movements.

Term structure of interest rates How interest rates on debt securities vary with respect to varying dates of maturity on the debt.

Theta Sensitivity of an option's price to changes in the option's time to expiry.

Tick The smallest unit of price change between successive transactions, 0.01 in some markets; 1/16, 1/8, or 1/4 in others.

Time weighted average return Geometric mean of daily returns accounting for fund inflows and outflows.

Trade A security transaction that creates a portfolio position based on an investment decision.

Trade-through The execution of an NMS stock trade at a price inferior to a protected quotation for that stock.

Trade-through rule See Order protection rule.

Trader Security market participant who trades, competing with others to generate profits, seeking compatible counterparties in trade, and seeking superior order placement and timing.

Trading arcade Sometimes referred to as a proprietary trading firm or prop shop, is a location for traders to work and trade from.

Trading floor broker NYSE member who work orders on behalf of clients.

Trading platform Computer system or software used to place or route quotes and transactions through a network to a financial intermediary or market.

Trading rebate Payment for order flow.

Transfer agent Processes orders to purchase and redeem a given security and maintains investor records.

Transparency The extent to which a market quickly disseminates high-quality information to the public.

Treasury Automated Auction Processing System The electronic system used by the U.S. Treasury (technically, the Fed) to take bids for its auctions of Treasury instruments.

Treasury bill Short-term zero coupon note issued by the U.S. Treasury. Considered to be relatively free of risk.

Treasury bond Long-term (2 to 30 years) coupon-bearing debt instrument issued by the U.S. Treasury. Considered to be relatively free of default risk.

Treasury note Intermediate-term (2 to 8 years) coupon-bearing debt instrument issued by the U.S. Treasury. Considered to be relatively free of default risk.

Treynor ratio Index of risk-adjusted portfolio performance, calculated as the portfolio's excess return to the risk-less rate relative to its beta.

Triangular arbitrage Currency arbitrage transactions intended to exploit relative price differences between one currency and two other currencies.

Turnover A measure of stock liquidity or activity calculated by dividing the total number of shares traded over a period by the average number of shares outstanding for that period.

UCITS See Undertakings for the Collective Investment in Transferable Securities.

Underlying asset Security or asset on which an option is written.

Undertakings for the Collective Investment in Transferable Securities (UCITS) (1) The regulatory framework allowing for the sale of cross-Europe mutual funds. (2) Funds qualified to trade under this regime.

Unit investment trust An entity that invests its funds in some portfolio of securities when it is established and does not normally rebalance its portfolio. The shares of the trust are often publicly traded.

Unlisted trading privileges (UTP) The right to trade a security on a particular exchange despite that security not meeting listing requirements of that exchange as authorized by the SEC and arising from the Unlisted Trading Privileges Act of 1994.

Upstairs markets Markets where institutions trade securities directly among themselves. Also called fourth markets.

UTP See Unlisted trading privileges.

Variance A statistical measure of dispersion, risk, and uncertainty. It is the expected value of the squared deviation of a data point from the expected value of the data set. It is the square of standard deviation.

Vega Sensitivity of an option's price to changes in the underlying security's standard deviation of returns.

Venture capital (VC) An alternative management vehicle that invests capital in private portfolio companies in development stages that have strong potential for rapid growth.

Vickrey auction See Second-price sealed-bid auction.

Vig (1) A spread that has been intentionally widened by a broker. (2) An interest rate charged by an illegal moneylender (loan shark).

Vigorish (1) An excessive interest rate on a loan. (2) An interest rate charged by an illegal moneylender (loan shark). (3) The percentage deducted from a gambler's winnings by the organizers of a game or gamble.

VIX The CBOE Volatility Index, a benchmark designed to gauge S&P 500 or stock market volatility. Based on the volatility implied by prices of option contracts written on S&P market indices. Also referred to as a fear index or fear gauge.

Volatility The unpredictable part of the stochastic process that is a measure of dispersion of the process from its predictable part.

Volatility index See VIX.

Volcker rule Limits banks and nonbank financial institutions supervised by the Fed with respect to proprietary trading and investment in and sponsorship of hedge funds and private equity funds.

Volume weighted average price (VWAP) The ratio of the total value traded to total volume traded over a given time horizon.

VWAP See volume weighted average price.

Walrasian auction A simultaneous auction where each buyer submits to the auctioneer his demand and each seller submits his supply for a given security at every possible price.

Wash sales Transactions intended to create the appearance of sales where, in effect, no sales actually take place.

Weak form efficiency Scenario where security prices reflect all information regarding historical prices.

Winner's curse Problem occurring when the winning bidder bids the most, and is the most likely to have bid too much for the auctioned object.

Yield curve Plotting of bond yields or market interest rates with respect to terms to maturity.

Yield to maturity The internal rate of return for a bond.

Zero cost collar A collar where the time zero proceeds from the call sale exactly offsets the cost of the put purchase. The options' exercise prices are set so that the call and the put have the same values.

Zero coupon bond A bond that makes no interest payments; it makes a single principal repayment at the bond's maturity.

Index

Note: Page numbers followed by "*f*," "*t*," and "*n*" refer to figures, tables, and notes, respectively.

A

Abel/Noser, 307
Abnormal return, 336, 363, 376, 380–381, 385, 389
Accredited investor, 107, 151
Acquisition of information, 5. *See also* Quotations/bids
ADRs. *See* American depositary receipts (ADRs)
Adverse selection, 169
  in dealer markets, 173–177
  defined, 171–172
  Glosten and Milgrom information asymmetry model, 179
  Kyle's model (traders, market makers, and noise traders), 173–177
    illustration, 175–176
    informed trader demand and dealer price adjustment, 174–175
  spread and, 177–179
Advisor, functions, 104
Agencies, 56–58
  credit rating, 61–63
Agency traders, 1–2
Algorithmic trading, 5–6, 113–115
  information algorithms, 114
  large transaction slicing (illustration), 114–115
  newsreader algorithms, 114
  strategies, 113–114
AllFirst Financial Corp., 430
Allied Irish Banks PLC (AIB), 430
Alpha, portfolio, 300–301
  Jensen, 302, 303*f*
    timing ability and, 303–306, 304*f*
Alpha models, 6, 113
Alternative trading system (ATS), 46–49
  crossing networks, 46
    internalization, 46
  dark pools, 46
  defined, 46
    by SEC, 46–47
  ECNs, 46
  sampling, 47*t*
  types, 46
  voice-brokered third-party matching, 46
Amaranth Advisors, LLC, 417
Amazon.com, 427
A.M. Best, 61, 148
American depositary receipts (ADRs), 45
American Exchange (AMEX), 36
  history of, 36
American Finance Association, 394
American options, 262
Ameritrade, 86
AMEX. *See* American Exchange (AMEX)
Amsterdam Bourse, 64
Amsterdam Stock Exchange, 38
Analyst estimates, for investors, 348
Analyst reports
  conflicts of interest and, 376–377
  information in, 376–377
Anchoring, 328
Annual reports, semistrong form efficiency and, 375–376
Anomalous efficiency
  prediction markets and, 391–394
Anti-Gold Futures Act of 1864, 136
ARA. *See* Arrow-Pratt absolute risk aversion coefficient (ARA)
Arbitrage, 4, 204–206
  classic, 178–179
  with currencies, 248–250
    PPP. *See* Purchase power parity (PPP)
    triangular arbitrage, 248–249, 249*t*
  with currency forward contracts, 250–255
    in FX markets, 250–255
  defined, 204
  with fixed income instruments and currencies, 227
  limitations on, 206–209
    basis risk, 206
    counterparty risk, 206–207
    implementation risk, 206
    model risk, 206
  in one-time period case, 267
  with options, 261
  pairs trading, 205–206
  PPP violation and, 252
  with riskless bonds, 227–229, 228*t*
  stat-arb (statistical arbitrage), 205–206
Arbitrage opportunity, 249. *See also* Arbitrage
Arbitrageur, 1–2, 204
Arcades, trading, 24
Archipelago Exchange, 51
Arrival price, 307, 309
Arrow-Pratt absolute risk aversion coefficient (ARA), 181

Ascending bid auction/English auction, 8
Asset value, 298–299
Assured Guaranty Municipal Corp., 58
AT&T, 331
ATM. *See* Automated teller machine (ATM)
ATS. *See* Alternative trading system (ATS)
Auctions, 7–12. *See also specific types*
  common value, 10–11
  Google's Dutch auction format, 11–12
  historical perspectives, 7
  outcomes, 10
  revenue equivalence theorem, 9–10
Automated teller machine (ATM), 66
Automated trading, 5, 113. *See also* Algorithmic trading
Axioms, von Neumann, 324–325
  cardinal utility function and, 324
    measurability/intermediate value, 324
    ranking, 324
    strong independence, 324
  investor rationality and, 324
    completeness/comparability, 324
    reflexivity, 324
    transitivity, 324

B

Bacanovic, Peter, 407
Back from the Brink: Coping with Stress, 429
Back office, 87
BaFin. *See* Federal Financial Supervisory Authority (BaFin)
Banging the close, 417–418
Bank for International Settlements, 66
Bank of America, 413
Bank of England, 162
Bank trusts, 107
Banker's acceptances, 59
Bankers' bourses, 64
Banking Acts of 1933. *See* Glass-Steagall Act
Barclays Capital, 43
Bargaining, 6–7
  relative power, 7
Barings Bank, 429–430
Base currency, 71
Basel Committee on Banking Supervision, 159–160
Basis risk, 206–207, 481
Batch market, 188–189
Bear Stearns, 149
Behavioral finance, 329–338
  decision process, simplifying, 337–338
  entertainment and, 342–343
  gender and, 341–342
  hyperbolic discounting, 334–336, 335*t*
  investor moods, weather, and investment returns, 336–337
  Monty Hall judgment error, 329–331
  myopic loss aversion, 333–334
  overconfidence. *See* Overconfidence of investors
  overconfidence and, 341–342
  overreaction, 336
  rational investors and diversification, 338
  similar ticker symbols and, 331
  underperforming funds, 332
  uninformed/irrational behavior, 332
Bell Atlantic, 331
Benchmark, 54, 64, 297–298, 306–307, 388, 422
Benchmarking
  portfolio performance, 300–301
    illustration, 302
  trade, 306
Bernoulli, Daniel, 322–323
Bernoulli, Nicholas, 322
Best execution, in institutional trading, 108–110
Betting, sports, 369
Bids. *See* Quotations/bids
Big Mac standard, 249
Bilateral auctions/double auctions, 9–10
Binomial option pricing model, 264. *See also* Black-Scholes options pricing model
  convergence to Black-Scholes model, 274, 275*t*
  illustration, 272–274
  multiplicative upward and downward movement values, obtaining, 271–272
  n time periods, extending to, 270–271
  three periods, extending to, 270–271
  two time periods, extending to, 268–269
Bitcoin, 72–74
Black box trading, 5, 113. *See also* Algorithmic trading
Black-Scholes options pricing model, 202–203, 215–219, 274–275. *See also* Binomial option pricing model
  applications, 275–276
  assumptions, 274–275
  binomial option pricing model convergence to, 274–281, 275*t*
  calls and puts, 215–219
  estimating implied volatilities, 221–225
  hedging in, 274–281
  sensitivities, 276–279
    delta, 276
    gamma, 277–278
    theta, 278
    vega, 278
Blockchain, 73
Bloomberg Professional, 23
Blue sky laws, 136–137
Bluffing, 413–423
Boesky, Ivan, 406–407

BofA Merrill, 43
Bond yields, risk, 229–231
  default/credit risk, 231
  interest rate risk, 231
  liquidity risk, 231
Bond(s), 3. *See also specific types*
  corporate, 59–61
  duration, 233–235
  indenture, 60
Bootstrapping, yield curve, 241–244, 242*t*, 243*f*
Bordeebook, 427
Boston Consulting Group, 101
Boston Options Exchange (BOX), 43
BOX. *See* Boston Options Exchange (BOX)
Brady Report, 164
Brick-and-mortar markets
  decline of, 49–53
  electronic vs. open outcry, 51–53
Broker(s), 1–2
  discount, 19
  full-service, 19
  NYSE, 36–37
  online, 19–20
  selection of, 19
Brokerage firms
  operations, 85–87
Brokered markets, 12
Brokers Act of 1696, 136
Brownian motion processes, 196–198, 197*f*
"Bruges Beurse," 36–38
Buffet, Warren, 370, 419
"Buttonwood Agreement," 38
Buy-side traders, 1–2. *See also* Institutional trading
  registered investment companies, 102–106
  unregistered investment companies, 106–108

C

C.A.S. Hewitt, 435
Call. *See* Call option
Call date, 62
Call market, 13–14
Call option, 3, 215–219. *See also* Options
Call price, 60
Callable bonds, 62
Canary Capital Partners, 412
Capital asset pricing model (CAPM), 300–301, 361–362
CAPM. *See* Capital asset pricing model (CAPM)
CAR. *See* Cumulative average residuals (CAR)
Cardinal utility function, 324
Castle Convertible Fund (CVF), 3
Catastrophe bonds, 60
CBOE. *See* Chicago Board Options Exchange (CBOE)
CBT. *See* Chicago Board of Trade (CBT)
CDOs. *See* Collateralized debt obligations (CDOs)
CDS. *See* Credit Default Swap (CDSs)
CEA. *See* Commodity Exchange Act (CEA)
Central counterparty, 54
Cerullo, Ed, 431
CFA. *See* Chartered Financial Analyst (CFA) Institute
CFTC. *See* Commodities Futures Trading Commission (CFTC)
Challenger space shuttle disaster, 391
Charles Ponzi Company, 434
Charles Schwab, 86
Chartered Financial Analyst (CFA) Institute, 306
Chase Medical Group, 419
Chiarella v. United States, 124, 371
Chicago Board of Trade (CBT), 36, 50, 69, 75
Chicago Board Options Exchange (CBOE), 36, 43–44, 51, 262
Chicago Daily News, 347
Chicago Mercantile Exchange (CME), 51, 71
Churning, 138
Circuit breakers, NYSE, 164–165
Citadel Investment Group, 118–119
Citadel Securities LLC, 43
CITIBLOC (Citigroup), 46
Classic arbitrage, 178–179
Clearance, trade, 5
Clearing firm, 54
Clearing house, 54
Clearing systems, 53–56
  trade comparison, 54–55
  trade settlement, 54–56
CLOB. *See* Consolidated limit order book (CLOB)
Closed-end investment company, 102–103
CME. *See* Chicago Mercantile Exchange (CME)
Co-location (proximity hosting), in HFT, 120–121
Coefficient of determination, 441, 444
Collateralized debt obligations (CDOs), 148
COMEX. *See* Commodities Exchange (COMEX)
Commercial paper, 60
Commodities, 3
Commodities Exchange (COMEX), 36
Commodities Futures Trading Commission (CFTC), 70, 75, 141, 154–155, 417
  background, 154–155
  operating unit, 155*n*9
Commodities trading regulation, updating of, 143
Commodity Exchange Act (CEA), 141, 155
The Commodity Exchange Act of 1936, 141
The Commodity Futures Modernization Act of 2000, 146
Commodity Futures Trading Commission Act of 1974, 143
Common stock, 2–3
Common value auctions, 10–11

Common value auctions (*Continued*)
symmetric information structure, 10–11
winner's curse problem, 11
Companies Act of 1844, 136
Companies Act of 1867, 136
Completeness/comparability, axioms, 324
Computer-initiated trades, 6
3Com, 208
Confirmation, of trade, 5
Conflicts of interest, analyst reports and, 376–377
Consolidated limit order book (CLOB), 143
Consolidated quotations system (CQS), 85, 125
Consolidated supervised entities (CSEs), 148
Consolidated Tape Association (CTA), 84
Consolidated trade system (CTS), 143
Consortium-based pools, 116
Constitution of the State of California, 136
Continuous double auction, 9–10
Continuous time process, 196
Convertible bonds, 60
Convexity, 237–239
Copeland and Galai options model, 182–183
Corners and pools, 138, 418–419
modern day corners, 419
Chase Medical Group, 419
Porsche, 419
Salomon Brothers, 419
Vanderbilt v. Drew, 418
Corporate and Criminal Fraud Accountability Act. *See* Sarbanes-Oxley Act of 2002 (SOX)
Corporate bonds and markets, 59–61
Corporate merger announcements, semistrong form efficiency and, 375–376
Corporate scandals, 146–149. *See also* Regulation(s)/regulatory system
Correlation coefficient, 441
Cortisol, trading profits and, 343
Counterparty risk, 78, 91, 207, 465, 481–482
Covariance, 440–441
Cowles, A., 373–374
Cox, Garfield, 373
CQS. *See* Consolidated quotations system (CQS)
Cramer, Gabriel, 322*n*1
CRE, 434–435
Credit default swap (CDSs), 77, 146, 432
Credit derivative product, 46, 57
Credit rating agencies, 63–65
The Credit Rating Agency Reform Act of 2006, 148
Credit Suisse Crossfinder, 46
Credit/default risk, 231
in forward exchange markets, 68
Crises, 142–146
1960s Go-Go Years and reform, 142
financial crisis of 2008, 146–149
Cross currency rates, 67, 70*t*
Crossing networks, 46. *See also* Alternative trading system (ATS)
defined, 53–55
internalization, 46, 54–55
upstairs markets and, 53–55
internalization, 54–55
CS First Boston, 11
CSEs. *See* Consolidated supervised entities (CSEs)
CTA. *See* Consolidated Tape (CTA)
CTS. *See* Consolidated trade system (CTS)
Cumulative average residuals (CAR), 374, 386–387, 387*t*, 388*t*
Currencies, 3–4
"long" position, 4
"short" position, 4
Currency exchange/foreign exchange
cross rates, 67, 67*t*
currency trading, 70
exchange rates, 66–67
forward exchange rates and contracts, 68–69
futures markets, 69–70
interdealer brokers and electronic broking, 70–71
markets, hedging in, 249–250
markets and, 66–72
orders, quotes, and spreads, 71–72
parity and arbitrage in, 249–250
PPP and, 250
Currency options
hedging exchange exposure with, 284–285
values, 244*t*
Currency trading, 70
Current yield, 230
Custodian, advisor as, 104
CVF. *See* Castle Convertible Fund (CVF)

D

Dark liquidity pools, 5, 46. *See also* Alternative trading system (ATS); crossing networks
institutional trading, 53
types, 116
Darley, John, 427–428
Darley's law, 427–428
DARPA. *See* Defense Advanced Research Projects Agency (DARPA)
Day order, 14
Day trading, 17–24
characteristics, 17
direct access trading, 19–20
online brokers, 19–20
platforms, 20–21
quotations and price data, 21–23, 22*f*, 23*f*
trading arcades, 24

DCF. *See* Discounted cash flow (DCF) analysis
De Montmort, Pierre Raymond, 322*n*1
Dealer markets, adverse selection in, 173–177. *See also* Kyle's adverse selection model
Dealer price adjustment/setting
  informed traders demand and, 190–193
  Kyle's adverse selection model, 188–189
Dealers, 1
Debentures, 60
Debt securities, 3, 55
  U.S. Treasury, 55–56
Decision process, simplifying, 337–338
Decision rule/test statistics, 442
Default/credit risk, 231
  in forward exchange markets, 68
Default waterfall, 96
Defense Advanced Research Projects Agency (DARPA), 393–394
Defined benefit plans, 106–107
Defined contribution plans, 106–107
Delphi, 62
Delta, 276
  hedging with, 280–281
Delta neutral portfolio, 277
Demsetz immediacy argument, 178
Demutualization, exchanges, 41–42
Depository Trust & Clearing Corporation (DTCC), 88–89
Depository Trust Corporation (DTC), 89
Depth, 14–17
Deregulation, 146–147. *See also* Regulation(s)/regulatory system
Derivative securities
  defined, 3, 261
Derivative securities markets
  hedging and, 261–262
  OTC markets and, 262
Descending bid auction/Dutch auction, 8–9, 9*t*
  Google's format, 11–12
Designated market makers (DMMs), 39–40
Diagonal matrix, 446–447
Diamonds (shares), 105–106. *See also* Dow Jones Industrial Average (DJIA)
Diebold Inc., 425
"Diminishing marginal utility," 323–324
Direct access trading systems, 19–20
Direct costs, financial regulatory activity and, 155–156
Direct market access (DMA)
  in HFT, 122
Direct participation programs issues (DPPs), 45
Discount broker, 19
Discount functions
  simultaneous estimation of, 244–245, 244*t*
Discounted cash flow (DCF) analysis, 377
Discrete time process, 196
Disposition effect. *See* Endowment effect
Distortions, to financial markets
  regulatory activity and, 156–158
Distributor, advisor as, 84
Diversification, rational investors and, 338
DJIA. *See* Dow Jones Industrial Average (DJIA)
DKs (don't knows), 88
DMA. *See* Direct market access (DMA)
DMMs. *See* Designated market makers (DMMs)
Dodd-Frank Wall Street Reform and Consumer Protection Act, 149–150, 167
Dollar-weighted return, 298–299
Dominion, 148
Double auctions/bilateral auctions, 9
Dow Jones, 23
Dow Jones Industrial Average (DJIA), 4, 105–106, 138, 365, 415–416
DPPs. *See* Direct participation programs issues (DPPs)
Drew, Daniel, 418
Drexel Burnham Lambert, 406
Dreyfus, 103
DTC. *See* Depository Trust Corporation (DTC)
DTCC. *See* Depository Trust & Clearing Corporation (DTCC)
Du Pont, 208
Duer, William, 406
Dummy variable approach to performance evaluation, 305–306
Dutch auction/descending bid auction, 8, 9*t*
  Google's format, 11–12

E

E-Signal, 23
E-Trade, 86
E.F. Hutton, 85
Earnings per share (EPS), 348
East Indies Trading Company, 64
EBS. *See* Electronic Broking Services (EBS)
ECNs. *See* Electronic communications networks (ECNs)
The Economist, 249
Edward I, King, 136
Effective half-spread, 109
Efficient market hypothesis (EMH), 373
Egan-Jones, 61
Eisen, Michael, 427
Electronic Broking Services (EBS), 71
Electronic communications networks (ECNs), 35, 46, 119–120
Electronic trading
  open outcry *vs.*, 51–53
Elementary statistics

Elementary statistics (*Continued*)
co-movement statistics, 440–441
hypothesis testing, 441–442
two populations, 442–443
mean, 439–440
overview, 439–446
simple OLS regression, 443–446
standard deviation, 439–440
variance, 439–440
EMH. *See* Efficient market hypothesis (EMH)
Employee Retirement and Security Act of 1974 (ERISA), 107, 142
EMTNs. *See* Euromedium-term notes (EMTNs)
Endowment effect, 328
English auction/ascending bid auction, 8
"Enron loophole," 146
Entertainment, overconfidence and, 309–310
EPS. *See* Earnings per share (EPS)
Equity portfolio, VaR calculation for, 314–315
Equity securities (stock), 3
ERISA. *See* Employee Retirement and Security Act of 1974 (ERISA)
ESMA. *See* European Securities and Markets Authority (ESMA)
ETFs. *See* Exchange-traded funds (ETFs)
Eurex Frankfurt, 71
Eurobonds, 64
Eurocredits, 64
Eurocurrency instruments and markets, 63–64
Eurodollar, 63, 72
Euromedium-term notes (EMTNs), 64
Euronext, 36, 38, 64–65
European options, 262
European Securities and Markets Authority (ESMA), 160
Event study methodology(ies), 379–389
market adjusted returns, 380
market and risk adjusted returns, 380
mean adjusted returns, 380
security prices and, 380
takeovers and, 381–389, 382*t*, 383*t*, 384*t*, 385*t*, 386*t*
Exchange options, 282–284
FX option valuation, 282–284, 283*t*
Exchange rates, 66–67
expectations, 254
forward exchange rates and contracts, 68–69
Exchange-based pools, 116
Exchange-traded funds (ETFs), 104
Exchange(s), 35. *See also* New York Stock Exchange (NYSE)
current NYSE equities membership types, 42–43
defined, 35
demutualization, 41–42
diminished floor trading activity, 42
early ancestors to modern, 36–37
floor markets, 35
governance changes, 41–42
national, 36, 37*t*
NYSE, transforming, 40–41
NYSE Euronext, 36
option, 43–45, 44*t*
NYSE circuit breakers, 164–165
NYSE rules, 74–75
privatization of rules, 162–165
regional, 36, 37*t*
as SROs, 35
traditional NYSE structure, 38–39
Execution, 4–5
Execution costs, 108–112
Expected utility maximization, 325, 329
Expected utility paradigm, 322–324
Expected value, of gamble, 322–323
Extreme value estimators, 201–202

F

F.I. Dupont, 142
Face value, 229–230
FactSet Research Systems, 23
Fat fingers, 425–427
FDIC. *See* Federal Deposit Insurance Corporation (FDIC)
"Fear of regret," 328
FED. *See* Federal Reserve Board (FED)
FED Regulation T, 18
Federal Deposit Insurance Corporation (FDIC), 86, 140
Federal Financial Supervisory Authority (BaFin), 162
Federal funds markets, 59
Federal funds rate, 59
Federal Home Loan Mortgage Corporation (FHLMC), 57–58
Federal Housing Administration (FHA), 57
Federal Ministry of Finance, 162
Federal National Mortgage Association (FNMA), 56–57
functions, 56–57
Federal Reserve Board (FED), 18, 146, 150
Fedwire system, 90, 92
"Feeder funds," 433
Feynman, Richard, 391
FGIC Financial Guaranty Insurance Company (FGIC)
FHA. *See* Federal Housing Administration (FHA)
FHLMC. *See* Federal Home Loan Mortgage Corporation (FHLMC)
Fidelity, 103
Fiduciary, 142, 144–145
Fiduciary standard, 141

Fill or kill orders, 15
Filter rules, market overreaction and, 363–364
Financial crisis of 2008, 146–149
Financial Guaranty Insurance Company (FGIC), 58
Financial Industry Regulatory Authority (FINRA), 18, 35, 45
Financial innovation, regulatory activities and, 157
Financial Institutions for Asbestos Reform, 378
Financial model-builders, 322
Financial Modernization Act of 1999. *See* Gramm-Leach-Bliley Act
Financial Services Authority (FSA), 162
Financial Stability Oversight Council (FSOC), 149
Financial statements, semistrong form efficiency and, 375–376
Financial Times, 331
Financialization, 27–29
FINRA. *See* Financial Industry Regulatory Authority (FINRA)
FINRA TRACE, 61
First-price sealed-bid auctions, 9
  winner's curse problem, 9
Fisher effect, 254
Fishing, 116–117, 420–421
Fitch, 58, 61–62, 148
Fixed income funds, 103, 231–232
Fixed income hedging, 229–233
  bond yields and sources of risk, 229–231
  fixed income portfolio dedication, 231–233
Fixed income portfolio immunization, 233–239
  bond duration, 233–235
  convexity, 237–239
Fixed-income securities and money markets, 55–64
  agency issues, 56–58
  corporate bonds and markets, 59–61
  credit ratings and credit agencies, 61–63
  Eurocurrency instruments and markets, 63–64
  money market issues of financial institution, 59
  municipal securities and markets, 58–59
  U.S. Treasury securities and markets, 55–56
Flash crash, 123, 158, 415–416, 426
Flash orders, 125
Flash trading, 125–128
  illustration, 126–127
Floating rate bonds, 63
Floor markets, 35. *See also* Exchange(s)
  diminished floor trading activity, 42
Floor trader, 24, 39
FNMA. *See* Federal National Mortgage Association (FNMA)
Football pool, 347, 347*t*
Foreign exchange (FOREX/FX). *See* Currency exchange/foreign exchange
Forte, Joseph S., 434
Forte, L.P., 434
Forward market
  exchange contract, 68–69
  transaction, 68–69
    risks in, 68–69
Forward market hedge, 256
Fourth markets, OTC, 46
Frames vs. substance
  prospect theory, 326–327
Fraud, 138
Friedman, Milton, 172
Friehling & Horowitz, 433
Friehling, David, 433
Frist, Bill, 377–378
Front running, 411–413
  parasitic trading and, 411–412
FSA. *See* Financial Services Authority (FSA)
FSOC. *See* Financial Stability Oversight Council (FSOC)
Full-service brokers, 18
Future Trading Act of 1921, 136
Futures contracts, 3, 261
Futures markets
  currency exchange, 69–70
FX. *See* Currency exchange/foreign exchange

G

Galleon Group, 408–409
Galway United Football Club, 429
"Gamer," 116–117, 420–421
Gamma, 277
Gamma neutral portfolios
  hedging with, 280–281
Gauss-Jordan Method, 449
Gender, overconfidence and, 341–342
General Motors, 208
General obligation bonds, 58
Geneva Trading, 24
GETCO Securities, 43
Geutfrend, John, 419
Giuliani, Rudolph, 407
Glass-Steagall Act, 140–141, 146
Global markets, 53
  regulations and, 135–136
Global Research Analyst Settlement, 377
Glore Forgan & Co., 142
Glosten and Milgrom information asymmetry model, 179
GNMA. *See* Government National Mortgage Association (GNMA)
Goldman Sachs, 41–43, 119, 148, 157–159, 408–409, 426
  Sigma X, 46

Goldman Sachs Tradable Unregistered Equity system (GSTrUE), 157
Good till canceled order, 14
Goodbody & Co., 142
Google
  Dutch auction format, 11–12
Gordan, Deborah, 349
Governance changes, exchanges and, 41–42
Government National Mortgage Association (GNMA), 57, 103
Government-Sponsored Enterprise issues, 56–58
Graham, Benjamin, 208
Gramm, Phil, 146
Gramm, Wendy, 146
Gramm-Leach-Bliley Act, 141, 146
Grasso, Richard, 41
Great Crash of 1929, 139
Great Depression, 57, 138–139
Greeks, Black-Scholes model. *See* Sensitivities, Black-Scholes options pricing model
GSTrUE. *See* Goldman Sachs Tradable Unregistered Equity system (GSTrUE)
Gumtech International (GUMM), 414–415

H
Haircut, 93
Half-spread, 109–110
Hall, Monty, 329–331
Hamilton, Alexander, 406
Hamilton, William Peter, 374
Hansen Transmissions, 427
Harlem Railroad Company, 418
Hedge funds, 107–108
Hedge ratio, 265–266
Hedgers, 1–2
Hedging
  in binomial environment, 264–274
  in Black-Scholes options pricing model, 274–281
  with currency forward contracts, 250–255
  with delta, 280–281
  derivative securities markets and, 261–262
  exchange exposure, 284–285
    with currency options, 284–285
    management, 255
  fixed income, 229–233. *See also* Fixed income hedging
  FX markets, 249–250
  with gamma neutral portfolios, 280–281
  with options, 261–262
  other strategies, 256–257
  two-time-period binomial options framework, 268–269
Herd behavior, 350
Hewett, Celia, 435
Hidden-size orders, 6
High-frequency trading (HFT), 33, 51, 118–125
  co-location (proximity hosting), 120
  latency arbitrage, 120–121
  strategies, 121–122
Hong Kong Futures Exchange, 52, 97
Hot potato volume, 426
House broker, 39
HRT Financial LLC, 43
Hunter, Brian, 417
Hussein, Saddam, 392
Hybrid markets, 13
Hybrids, 261
Hyperbolic discounting, 334–336, 335*t*
Hypothecation agreements, 18
Hypothesis testing, 441–442

I
IBES. *See* Institutional Brokers Estimate System (IBES)
Iceberg, 6, 116
Identity/unit matrix, 447
IEX. *See* Investors Exchange
Illegal insider trading, 404–411
  defined, 144
ImClone, 407–408
IMM. *See* International Monetary Market (IMM)
Immediacy argument (Demsetz), 178
Immediate or cancel orders, 15
Implementation shortfall, 307, 309–311
  illustration, 310–311, 311*t*
Implied volatilities, 202–204, 224, 278
  estimating, 202–203
    aggregating procedures, 225
    simple closed-form procedures, 224–225
Independent broker, 39
Index arbitrage, 164
Index funds, 103, 106
Indexed bonds, 60
Indices, 4
Indirect costs, financial regulatory activity and, 156
"Industry standard," 329
InfoReach, 122
Information
  acquisition of, 5
  asymmetries of, 178–179, 185
  economics of, 171
  in publications and analyst reports, 376–377
  trading and, 171–172, 178
Information asymmetry model (Glosten and Milgrom), 179
Information cascade, 349–350
Informed traders, 173–177. *See also* Kyle's adverse selection model

Initial public offering (IPO) anomaly, 367
Insider trading, 140
  illegal, 144, 404–411
  monitoring inside trading activity, 410–411
  notorious cases, 405–410
    Albert H. Wiggin, 406
    Dennis Levine, Ivan Boesky, and Michael Milken, 406–407
    Martha Stewart, 407–408
    Raj Rajaratnam, 408–409
    William Duer, 403
  regulation, 143–144
  strong form efficiency and, 373–379
Insider Trading and Fraud Act of 1988, 405
Insider Trading Sanctions Act of 1984, 405
Instinet, 49
Institutional Brokers Estimate System (IBES), 355
Institutional trading, 105
  algorithmic trading, 6, 117–119
  best execution, execution costs, and price improvement, 112–116
  buy-side institutions
    registered investment companies, 106–110
    unregistered investment companies, 110–112
  dark pools, 48, 119–121
  flash trading and sponsored access, 129–132
  high-frequency trading, 122–129
  market impact, 15, 105–106
  stealth and sunshine trading, 121–122
Institutions Institutional trading market impact and, 105–106
Instruments
  categories, 1–5
  securities and, 1–5
Interactive Brokers, 20–21, 89
Interdealer brokers, 72–73
Interest Equalization Tax, 65–66
Interest rate parity (IRP), 257
  violations, arbitrage and, 255
Interest rate risk, 235
Intermarket, 88, 96
Intermarket trading system (ITS), 88, 96, 129
Internal Revenue Service (IRS), 18–19
Internalization, 56–57
Internalization pools, 120
International equilibrium model, 255
International Fisher effect, 258–259
International Monetary Market (IMM), 73
International Organization of Securities Commissions (IOSCO), 163
International Securities Exchange (ISE), 38, 45
International Swaps and Derivatives Association (ISDA), 78, 167
Internet, 24, 45, 51–52
Intrade, 392–393
Inventory model (Stoll), 184–186
Inverting matrices, 449–450
Investing, defined, 1
The Investment Advisers Act of 1940, 145
The Investment Companies Act of 1940, 145
Investment company(ies). *See also* Institutional trading
  closed-end, 106–107
  managed
    fees and structures, 107–110
    types, 106–108
  mutual funds. *See* Mutual funds
  open-end, 106–107
  performance evaluation, 305–306
  registered. *See* Registered investment companies
  unregistered. *See* Unregistered investment companies
Investment portfolio performance, evaluation. *See* Portfolio performance, evaluation
Investment returns
  investor moods and, 341–342
  weather and, 341–342
Investor(s), 267–268, 302, 306
  behavioral finance. *See* Behavioral finance
  consensus opinion, 351–355
    analyst estimates, 353
    football predictions, 352, 352*t*
    herds and swarms, 354–355
    rational investors and price setting, 343
  irrational and rational investors, 336–337
  neurofinance, 349–350
  prospect theory. *See* Prospect theory
  rational investor paradigms. *See* Rational investors
Investors Exchange, 39*t*, 50*t*
IOSCO. *See* International Organization of Securities Commissions (IOSCO)
Iowa Electronic Markets, 392–393
IPO. *See* Initial public offering (IPO) anomaly
IPO 2004, 12–13
IQFeed, 24–25
IRP. *See* Interest rate parity (IRP)
Irrational investors, 334, 336, 351, 353
IRS. *See* Internal Revenue Service (IRS)
ISDA. *See* International Swaps and Derivatives Association (ISDA)
ISE. *See* International Securities Exchange (ISE)
ITS. *See* Intermarket trading system (ITS)

J

J-com shares, 432
J.D. Power and Associates, 89
January effect, 371–372

January effect (*Continued*)
tax-driven selling and, 371–372
window dressing and, 372
Japanese Financial Services Agency, 166
Javitz, Jacob, 146–147
Jensen measure, 305–306
alpha, 307, 307*t*
timing ability and, 308–309, 309*f*
Jett, Orlando Joseph, 437–438
Jitney game, 420
JPMorgan, 422, 432
Junk bonds, 64

K
Kahneman, Daniel, 330–331
Kansas City Board of Trade, 38
KCG, LLC, 123
Kerviel, Jérôme, 438–439
Kidder Peabody, 437–438
Knight/Trimark Group, 55
Kyle's adverse selection model, 177–178
illustration, 179–180
Kyle's lambda, 179, 193

L
Langone, Kenneth, 43–44
Late trading, 418–420
market timing and, 308–311
Latency, 13, 123–124
Latency arbitrage, 124–125
Law of one price, 208–209, 253, 297
Lawrence, Peter, 427
Lay, Kenneth, 389
Layering, 414
Lebed, Greg, 417
Lebed, Jonathan, 417
Leeson, Nick, 436–437
Lehman Brothers, 80, 94, 152–153, 419
Levine, Dennis, 413–414
LIBOR. *See* London Interbank Offered Rate (LIBOR)
LIFFE. *See* London International Financial Futures Exchange (LIFFE)
LIFFE CONNECT, 54
Limit order, 15
Limited obligation bonds, 58
Limit up-Limit down, 164–165
Liquidity, 15–18
defined, 15–16
dimensions, 16
market depth, 16
slippage, 16
width/tightness, 16
Liquidity rebate traders, 123. *See also* High-frequency trading (HFT)
Liquidity risk, 71, 235
in forward exchange markets, 70
Liquidnet, 7–8, 48, 56
LLC, 43, 45, 53, 63, 91
Lo, Andrew, 343
Load funds, 107–108
Log-utility function, 328–329
Logit modeling, 64
London Bullion Market Association fixing process, 9
London Interbank Offered Rate (LIBOR), 66
London International Financial Futures Exchange (LIFFE), 54, 73
London Stock Exchange, 40, 140
London stock market, 66–67
Long Term Capital Management (LTCM), 211, 213
LTCM. *See* Long Term Capital Management (LTCM)

M
M-square measure, 301
Macaulay simple duration formula, 238
MacLeod, Bentley, 343
Madoff, Bernard, 55, 431
Madoff Securities, 432–433
Major Depression-era legislation, 140–142
Make-or-take pricing, 424
The Making of a Fly: The Genetics of Animal Design, 427
Maloney Act of 1938, 145, 167
Margin, 4, 69, 75, 92–93, 462
Market adjusted returns
event study methodologies and, 386
Market and risk adjusted returns
event study methodologies and, 386
Market depth, 16, 17*t*
Market efficiency, 365
anomalous efficiency, 391–394
event study methodology. *See* Event study methodology(ies)
overview, 359
prediction markets, 391–394
random walks and, 195–196
Brownian motion processes, 196–198, 197*f*
martingales, 195–196
submartingales, 196
Wiener processes, 196–197
semistrong form efficiency. *See* Semistrong form efficiency
strong form efficiency and insider trading, 389–391
testing momentum and mean reversion strategies, 370–373, 371*t*, 372*t*
weak form efficiency. *See* Weak form efficiency

Market impact, 14, 101–102, 109–110, 130
  institutions and, 101–102
Market makers, 173–177. *See also* Kyle's adverse selection model
Market manipulation, 413–423
  "pump-and-dump" schemes, 416
  banging the close, 417–418
  corners and pools. *See* Corners and pools
  fishing, 420–421
  posting quotes, 423
  predatory algo trading, 421
  quote matching, 423
  quote stuffing, 423
  wash sales, 138, 420
Market microstructure, 12–14, 169–170
  execution systems, 12–13
Market order, 14
Market overreaction
  filter rules and, 363–364
Market Participation Identification Numbers (MPIDs), 127
Market performance, 337–338
Market timing
  late trading and, 411–413
  selection *vs*., investment portfolios and, 303–306
    dummy variable approach to evaluating, 305–306
    quadratic variable approach to evaluating, 303–305
Market volatility, semistrong form efficiency and, 378–379
Market-on-close (MOC), 15, 307
Market(s)
  currency exchange and, 69–72. *See also* Currency exchange/foreign exchange
  fixed-income securities and. *See* Fixed-income securities and money markets
  fragmented, 49
  global, 53, 65*t*, 91, 163
  historical perspectives, 66
  opaque, 5
  sports betting, 369
  transparent, 5
"Marking to the market," 69, 75
Markov process. *See* Random walks
Martha Stewart Living Omnimedia, Inc., 407–408
Martingale process, 196
Massmutual Corporate Investors (MCI), 331
Matching index, 103
MATIF, 52
Matrix arithmetic, 446–451
Matrix/matrices
  defined, 446
  diagonal matrix, 446–447
  identity/unit, 447
  inverting, 449–450
  multiplication of, 448–449
  principal diagonal, 446–447
  scalar, 446
  solving systems of equations, 450–451
  square matrix, 446
  symmetric matrix, 446–447
  vector, 446
Matter under inquiry, 153
Maverick Trading, 24
"May Day," 143
MBSs. *See* Mortgage-backed securities (MBSs)
McGregor, Ewan, 429
MCI. *See* Massmutual Corporate Investors (MCI)
MCI Communications (MCIC), 331
MCIC. *See* MCI Communications (MCIC)
Mean, 439–440
Mean adjusted returns
  event study methodologies and, 380
Mean reversion
  in stock prices, 362
  strategies, testing momentum and, 370–373, 370*t*, 372*t*
Measurability/intermediate value, 324
Meriwether, John, 207, 419
Merrill Lynch, 85–86, 104, 137, 148, 406
Merton, Robert, 207
MetaStock, 23
MidPoint Match, 116
Midwest Exchange, 36
MiFID, 160–161
Milken, Michael, 406–407
Miller, Edward, Dr., 336
Mizuho Financial Group, 425
Momentum
  in stock prices, 361
  testing, mean reversion strategies and, 370–373, 370*t*, 372*t*
Money market hedge, 256
Money Market Issues of Financial Institutions, 59
Money market mutual funds, 59
Monty Hall judgment error, 329–331
Moody's, 61–63, 62*t*
Morgan Stanley, 11, 65, 148, 425
Morgenstern, Oscar, 324–325
Mortality frame, 327
Mortgage-backed securities (MBSs), 9, 57, 62, 148
Morton Thiokol, 391
Moving averages, 364–365
Mozer, Paul, 419
MPIDs. *See* Market Participation Identification Numbers (MPIDs)

MS Pool (Morgan Stanley), 46
Mulheren, John A., 407
Multidiscriminate analysis, 62
Municipal bond funds, 103
Municipal securities markets, 58
Muskie, Edmund, 133, 142
Mutual funds, 102–103
  fees and structures, 103–104
Myopic loss aversion, 333

N
Naked access, 127
NASD. *See* National Association of Securities Dealers (NASD)
NASDAQ, 7, 20–22, 23*f*, 38, 41, 54, 71
  demutualization, 41
NASDAQ Crossing, 54
NASDAQ OMX, 43, 71, 390
NASDAQ Stock Market, Inc., 37*t*
National Association of Securities Dealers (NASD), 38, 45, 141
National Association of Security Dealers Automated Quotation System (NASDAQ). *See* NASDAQ
National best bid or offer (NBBO), 42, 116, 124–125, 148
National exchanges, 36, 37*t*. *See also* Exchange(s)
National Futures Association (NFA), 75, 143, 163
National market system (NMS), 84, 143
  updating, 143
National Securities Clearing Corporation (NSCC), 88–91
Nationally Recognized Statistical Rating Organizations (NRSROs), 61
NAV. *See* Net asset value (NAV)
NBBO. *See* National best bid or offer (NBBO)
Negotiable certificates of deposit, 59
Net asset value (NAV), 103–104
  computing, 298–299
  illustration, 299–300, 299*t*
Netting, 89
Neurofinance, 344–345
New York Curb Exchange, 38
New York Mercantile Exchange, 3, 50–51, 69, 75
New York Stock Exchange (NYSE), 5–6, 12–13, 17, 36, 54, 64–65. *See also* Exchange(s)
  circuit breakers, 164–165
  current equities membership types, 42–43
    DMM, 42–43
    SLP, 42–43
    trading floor broker, 42–43
  demutualization, 41–42
  diminished floor trading activity, 42
  floor trader, 39
  history of, 36
  house broker, 39
  independent broker, 39
  listing standards, 163
  options markets, 43, 44*t*
  specialist, 39
  traditional structure, 38–39
  transforming, 40–41
New York Stock Exchange Group, 36
Newton-Raphson method, 203, 222, 224*t*, 278
NFA. *See* National Futures Association (NFA)
1960s Go-Go Years and reform, 142
NMS. *See* National market system (NMS)
No-load funds, 103–104
Noel, Walter, 433
Noise traders, 172. *See also* Kyle's adverse selection model
Normal returns
  event study methodologies and, 380
Not held order, 14
Novation, 88, 91, 99
NRSROs. *See* Nationally Recognized Statistical Rating Organizations (NRSROs)
NSCC. *See* National Securities Clearing Corporation (NSCC)
NSDQ. *See* NASDAQ
NYSE. *See* New York Stock Exchange (NYSE)
NYSE Amex, 43
NYSE Arca, 43
"NYSE Bonds," 60–61
NYSE Regulation, Inc., 41
NYSE Rule 80A, 164
NYSE Rule 80B, 164

O
OCC. *See* Options Clearing Corporation (OCC)
Odd-Eighths scandal, 151, 421
Offers, 1–2
Oliver Wyman, 158–159
OLS. *See* Ordinary least square (OLS) regression
Online brokers, 19–20
Opaque markets, 5
Open outcry trading
  electronic vs., 51–53
Open-end investment companies, 102–103
Oppenheimer, 19
Option exchanges, 43–45, 44*t*. *See also* Exchange(s)
  competition among, 43–44
  technological advances, 44–45
Options, 3, 43–44, 202, 215–221, 261
  American, 262
  arbitrage and hedging with, 261
  in binomial environment, 264–274

calls, 3, 203, 215–219, 263
European, 264, 267
puts, 3, 215–219, 263
Options Clearing Corporation (OCC), 43, 88, 97
Options Price Reporting Authority, 85
Order Book, 21, 40, 78, 124
Order precedence rules, 16–17
Order processing costs
institutional trading, 101, 108
Order Protection Rule, 148, 158, 168
Order-driven markets, 12–13
Orders, 14–17, 71–72
types, 14–15
Ordinary least square (OLS) regression, 110–111, 371, 443
simple, 443–446
Ornstein-Uhlenbeck function, 247
Ossie, James, 434
OTC markets. *See* Over-the-counter (OTC) markets
OTC Markets Group, Inc., 45–46
OTCBB. *See* Over-the-Counter Bulletin Board (OTCBB)
Out-trades, 88–89
Outcomes, auctions, 10
Over-the-counter (OTC) markets, 27, 45–49
defined, 45
derivative securities and, 262
fourth markets, 46
third markets, 46
upstairs markets, 46
venues, 45
Over-the-Counter Bulletin Board (OTCBB), 45
Overconfidence of investors, 339–342, 340*t*
Overreaction
of investors, 332–336
market, filter rules and, 363–364

P

Pacific Exchange, 36, 51
Paine Webber, 85, 431
Pairs trading, 205–206
Parasitic trading
front running and, 411–412
Paris Bourse, 64
Paris Exchange, 64
Parity, 249–250
Parkinson estimator, 201–202
Parmalat, 62
Pass-through securities, 57–58
Passive market-making, 121
Pattern day trader, 18, 140
Payment for order flow, 423–424
PBOT. *See* Philadelphia Board of Trade (PBOT)
PCAOB. *See* Public Company Accounting Oversight Board (PCAOB)
Pegging, 114
Penny-jumping, 412
Pension funds, 106–107
Perfect foresight half spread, 110–112
slippage and, 110–112, 111*t*, 112*t*
transaction size and, 110–112, 111*t*, 112*t*
Pershing, 88
PFSWeb, 208
Phantom trading profits, 431
Philadelphia Board of Trade (PBOT), 71
Philadelphia Exchange (PHLX), 43
Philadelphia-Baltimore-Washington Exchange, 36
PHLX. *See* Philadelphia Exchange (PHLX)
Pickard & Co., 142
Ping destinations, 116
"Pink Sheets," 45
Poindexter, John, 394
Political intelligence units, 377–378
Polynomials, derivatives of, 451–453
Pomper, Alexander, 414–415
Ponzi, Carlo, 433–434
Ponzi schemes
trading and, 432–435
other schemes, 434–435
Ponzi, Carlo, 433–434
Pools. *See* Corners and pools
Porsche, 419
Portfolio, fixed income
dedication, 231–233
immunization, 233–239
Portfolio immunization, 235–237
Portfolio performance, evaluation, 297–303
alpha measure. *See* Alpha, portfolio
benchmarking, 300–301
illustration, 302–303
market performance and, 302*t*, 305*t*
NAV and returns
computing, 298–299
illustration, 299–300, 299*t*
standard deviation, 300–301
Posit Marketplace (ITG), 54
Posting quotes, 423
Power of a test, 442
PPP. *See* Purchase power parity (PPP)
Pre-1930s securities regulation, 136–138
Predatory algo trading, 421
Prediction markets, 392–394
anomalous efficiency and, 391–394
Preferred stock, 3
Prerouting displays, 125
Present value, 230, 233
Price data, quotations and, 21–23, 22*f*, 23*f*
Price improvement

Price improvement (*Continued*)
institutional trading, 108–112
Price multiples, 377
Price priority, 16–17
Price setting
rational investors and, 348–349
Price/earnings effects, small firms and, 366–367
Primary markets, 55
Principal diagonal, 446
Principles-based regulatory approach, 168
Private banking, 107
Private bourse, 64
Private equity, 107
Private equity firms, 106–107
Privatization
defined, 58
NYSE circuit breakers, 164–165
NYSE rules, 164
of regulation and exchange rules, 162–165
Probability space, 196
Probit model, 62
Profnath, 427
Prop shop, 24
Proprietary traders, 1–2
Prospect theory, 325–328
anchoring, 328
endowment effect and, 328
frames vs. substance, 326–327
losses and inconsistency, 325–329
Prudent man rule, 142
Public bourse, 64
Public Company Accounting Oversight Board (PCAOB), 147
Public crossing networks, 116
Publications, information in, 376–377
"Pump-and-dump" schemes, 416
Purchase power parity (PPP), 248–250
in forward markets, 249
FX rates, 249
relative PPP condition, 251
in spot markets, 251–252
violation, arbitrage and, 252
Pure discount notes, 241, 244
Pure expectations theory, 240, 243
Put sensitivities, 281
Put stock option, 3, 215–219, 263
Put-call parity, 263–264
Put-call relation, 263

Q

QIBs. *See* Qualified institutional buyers (QIBs)
Qualified institutional buyers (QIBs), 151
Quasi-arbitrage, 205
Questrade, 20
Quick & Reilly, 86
Quotations/bids, 1–2, 21–23, 54, 71–72, 84–85
conventions, 72
in currency trading, 70–71
price data and, 21–23, 22*f*, 23*f*
Quote matching, 186
Quote Rule (SEC Rule 602 of Regulation NMS), 125
Quote stuffing, 414
Quote-driven markets, 12
Quoted spread, 109

R

Racketeer Influenced and Corrupt Organizations Act (RICO), 407
Rajaratnam, Raj, 408–409
Random walks, 361
market efficiency and, 195–198
Brownian motion processes, 196–198
martingales, 195–196
Wiener processes, 196, 198
Ranking, 324
Rate risk, in forward exchange markets, 68
Rational investors, 321–325
axioms and. *See* Axioms, von Neumann
diversification and, 338
expected utility, 322–324
irrational and, 346
price setting and, 348–349
St. Petersburg paradox, 322–324
Von Neuman and Morgenstern axioms, 324–325
Raymond James, 19, 86
Reagan, Ronald, 146
Realized half-spread, 109
REDIBook, 47
Reflexivity axiom, 324
Regional exchanges, 36, 37*t*. *See also* Exchange(s)
Registered investment companies, 102–106. *See also* Institutional trading
ETFs, 104–106
managed investment company, 102–104
Regressions, 443
OLS. *See* Ordinary least square (OLS) regression
Regulation M (reserve requirements), 63
Regulation NMS, 125, 148, 158
Regulation Q (interest ceilings), 63
Regulation T, 18
Regulation(s)/regulatory system, 133
activities, impacts of, 155–156
CFTC. *See* Commodities Futures Trading Commission (CFTC)
commodities trading regulation, 143

corporate scandals, 146–149
crises, 142–146
1960s Go-Go Years and reform, 142
financial crisis of 2008, 146–149
deregulation, 146–147
Dodd-Frank Wall Street Reform and Consumer Protection Act, 149–150
insider trading regulation, 143–144
National Market System (NMS), 143
post-SOX, 147–148
principles-based approach to regulation, 134
privatization of, 162–165
purpose of, 133
regulatory approaches and regulatory balance, 133–136
pre-1930s securities regulation, 136–138
regulatory approaches, 134–135
rules-based approach to regulation, 134
Sarbanes-Oxley, 146
SEC. *See* Securities and Exchange Commission (SEC)
self-regulation, government oversight of, 151–155
U.S. securities market legislation, 139–142
additional major depression-era legislation, 140–142
The Securities Act of 1933, 139–140
The Securities Exchange Act of 1934, 139–140
Relative strength rules, 363
Remote Market Makers (RMMs), 45
Renaissance Technologies LLC, 119
Repurchase agreements (repos), 59
Retek, 208
Return variance, 198–199
and spreadsheet applications, 215, 216*t*
Returns, for portfolio
computing, 298–299
dollar-weighted, 298–299
illustration, 299–300, 299*t*
time-weighted average, 299
Revenue equivalence theorem, 10–11
Risk, 198–204
basic measures, 203, 203*t*
defined, 198–199
extreme value estimators, 201–202
historical volatility indicators, 199–201, 200*t*
implied volatilities, 202–204
portfolio, measurement, 300
return variance, 198–199
and spreadsheet applications, 215, 216*t*
variance, 198–199
RMMs. *See* Remote Market Makers (RMMs)
Robotrading, 5, 113. *See also* Algorithmic trading
Rogue Trader, 431
Rogue traders, 428–432
Jérôme Kerviel and Société Générale, 431–432
John Rusnak and AllFirst, 430
Nick Leeson and Barings Bank, 429–430
Orlando Joseph Jett and Kidder Peabody, 430–431
Rogue trading, 427–432
Roll, Richard, 370, 394
Roosevelt, T., 139
Rosenthal, Alan, 407
Rothschild, Nathan, 404
Routing, 5, 40–41
Royal Dutch/Shell Group, 207–208
Rule 10b-5 of Securities Exchange Act, 144
Rule 144A markets, 151
Rule 300(a) of SEC Regulation ATS, 47–48
Rules-based regulatory approach, 134
Rusnak, John M., 430

S

S&P 500 Index, 4
S&P Depository Receipt (SPDR), 104
Salomon Brothers, Inc., 207, 419
Sarbanes-Oxley Act of 2002 (SOX), 147, 156
benefits and costs associated with, 157
post-SOX, 147–148
Scalar, 446
Scholes, Myron, 207
ScotTrade, 19
SDBK. *See* Super display book system (SDBK)
SEC. *See* Securities and Exchange Commission (SEC)
SEC Rule 602 of Regulation NMS (Quote Rule), 125
Second-price sealed-bid auction (Vickrey auction), 9
Secondary markets, 55
Section 10(b) of Securities Exchange Act, 144, 167, 405
Section 14(e)-3 of Securities Exchange Act, 144, 405
Secured overnight financing rate (SOFR), 64, 77, 422
Securities, 1–4. *See also specific types*
instruments, 2–4
overview, 1–4
trading, 4–6
Securities Act of 1933, 139–140
Securities Act of 1975, 84
Securities Acts Amendments of 1975, 143, 148
Securities and Exchange Act of 1934, 35, 152, 406
Securities and Exchange Commission (SEC), 18, 44, 101–102, 152–153, 377
Rule 144A, 151
Securities and Exchange Surveillance Commission (SESC), 162
Securities Exchange Act of 1934, 49, 84, 139–140, 405
Securities Industry and Financial Markets Association, 158
Securities Industry Automated Corporation (SIAC), 84, 88

Securities Investor Protection Act, 86, 133, 142
Securities Investors Protection Corporation (SIPC), 86
Securities markets, 35
  basic components, 12
  day trading, 17–19
  execution systems, 12
  microstructure, 12–14
  orders, liquidity, and depth, 14–17
Securities Observation, News Analysis and Regulation (SONAR) system, 411
Securities trading, 4–6
  algorithmic, 5
  auctions. *See* Auctions
  bargaining. *See* Bargaining
Self-regulatory organizations (SROs), 35, 152, 163. *See also* Regulation(s)/regulatory system
Sell side traders, 2. *See also* Institutional trading
Semistrong form efficiency, 373–379
  corporate merger announcements, annual reports, and other financial statements, 375–376
  DCF analysis and price multiples, 377
  early tests, 373–374
  event studies, 379–389
  information in publications and analyst reports, 376
  market volatility, 378–379
  political intelligence units, 377–378
  stock splits, 374–375
Sensation seeking, overconfidence and, 341–342
Sensitivities, Black-Scholes options pricing model, 276–281
  calculations for calls, 279–280
  delta, 276
  gamma, 277
  theta, 278
  vega, 278
Separate Trading of Registered Interest and Principal Securities (STRIPS) program, 56
Serial bonds, 60
Series EE U. S. Savings Bonds, 56
Series H U. S. Savings Bonds, 56
Series I U. S. Savings Bonds, 56
SESC. *See* Securities and Exchange Surveillance Commission (SESC)
Settlement, of trade, 5, 89–90
1720 Bubble Act, 136
Sharpe ratios, 300–301
Short S&P 500 ProShares, 105
Short sell, 6, 208, 406, 418
SIAC. *See* Securities Industry Automated Corporation (SIAC)
Siegel, Jeremy, 333
Siegel, Martin, 406
Simple time slicing, 114
Sinking fund, 60
SIPC. *See* Securities Investors Protection Corporation (SIPC)
Slippage, 15, 20, 53, 109, 113. *See also* Market impact
  perfect foresight half spread and, 110–111, 111*t*, 112*t*
  transaction size and, 110–111, 111*t*, 112*t*
SLM Corporation, 58
SLMA. *See* SLM Corporation
SLPs. *See* Supplemental liquidity providers (SLPs)
Small firms
  price/earnings effects and, 366–367
Smart order routing, 114
Smith Barney, 85, 137, 419
"social interaction barometer," 349
Société Générale, 431–432
Software-based trading platforms, 20
SONAR. *See* Securities Observation, News Analysis and Regulation (SONAR) system
Sortino coefficient, 301
SOX. *See* Sarbanes-Oxley Act of 2002 (SOX)
SPDR. *See* S&P Depository Receipt (SPDR)
Spear, Leeds & Kellogg Specialists LLC, 43
Specialist (NYSE), 39
Speculators, 1–2, 4
Spider fund, 104
Spitzer, Elliot, 137, 377, 412
"split strike conversion," 433
Sponsored access, 125–128
Spoofing, 414–415
  illustration, 414–415
Sports betting markets, 369
  sports forecasting, 347
Spot markets, PPP in, 250
Spot rate, 241
Spread(s), 71–72, 109–110, 169–170
  adverse selection and, 177–178
  Copeland and Galai options model, 182–183
  Demsetz immediacy argument, 178
  Glosten and Milgrom information asymmetry model, 179
Square matrix, 446
SROs. *See* Self-regulatory organizations (SROs)
St. Petersburg paradox, 322–324
Standard & Poor's, 61, 62*t*, 148
Standard deviation, 215, 221–222, 439–440
  portfolio, 300–301
Standard Portfolio Analysis of Risk (SPAN), 93, 313
Stanford, R. Allen, 435
Stanford International Bank, 435
Stat-arb (statistical arbitrage) strategies, 205–206
State space, 196
State Street Bank, 104

Steadman Technology & Growth Fund, 332
Stealth trading, 117–118
Stern, Edward Julius, 412
Stern, Leonard N., 412
Stewart, Martha, 407–408
Stiel, Benn, 159
Stochastic process, 196
STOCK Act of 2012, 144, 378, 390, 410
Stock funds, 103
Stock option, 188. *See also* Options
Stock splits, 374–375
Stockholm Exchange, 41
Stocks for the Long Run, 333
Stoll inventory model, 180–182
Stop order, 14
Stratos Lightwave, 208
Striking price/exercise price, 215
STRIPS program. *See* Separate Trading of Registered Interest and Principal Securities (STRIPS) program
Strong, Richard, 413
Strong form efficiency
  insider trading and, 389–390
Strong independence, axioms, 324
Strong Mutual Funds, 413
Studebaker Corporation, 142
Student Loan Marketing Association (SLMA). *See* SLM Corporation
Submartingales
  defined, 196
Substance vs. frames
  prospect theory, 326–327
Sunshine trading, 117–118
Super designated order turnaround system (SuperDOT), 40
Super display book system (SDBK), 40
SuperDOT system. *See* Super designated order turnaround system (SuperDOT)
SuperMontage, 44
Supplemental liquidity providers (SLPs), 42
Survival frame, 326–327
Swap Execution Facility (SEF), 78, 150
Swaps, 3, 76–79, 261
Swarms. *See* Herd behavior
Sydney Futures Exchange, 52
Symmetric matrix, 446–447

T

T. Rowe Price, 103
Tailgating, 412
Takeovers, event studies and, 381–389, 382*t*, 383*t*, 384*t*, 385*t*, 386*t*, 387*t*, 388*t*
Tâtonnement, 8
Tax loss write-offs, 328
Tax-driven selling, January effect and, 365–366
TD Ameritrade, 19
Tele-Communications Inc., 331
Term structure, 239–248
  estimation with coupon bonds, 241
  of interest rates, 240–241
Terrorism Information Awareness Program (TIA), 394
Test statistics/decision rule, 442
Testosterone, trading profits and, 343
TFX. *See* Tokyo Financial Exchange (TFX)
Thain, John, 41
Theta, 278
Thinkorswim, 20
Third markets, OTC, 46
Thomson Reuters, 23, 422
TIAA-CREF, 338
Time Magazine, 392
Time priority, 16–17
Time-weighted average return, 299
Tokyo Financial Exchange (TFX), 71
Tokyo Stock Exchange, 425
Touting, shares of stock, 416
Trade action, 1
Trade allocation, 5
Trade benchmarking, 306
  implementation shortfall. *See* Implementation shortfall
  VWAP and. *See* Volume-weighted average price (VWAP)
Trade decisions, 1
Trade(s)
  comparison, in clearing process, 87
  overview, 1–4
  settlement, in clearing process, 55–56
Tradeking, 19
Traders
  buy side, 1–2
  overview, 1–4
  sell side, 2
Tradesports.com, 394
TradeStation, 20
Trading, 1–2, 406
  adverse selection and. *See* Adverse selection
  algo program, 427
  algorithmic. *See* Algorithmic trading
  bargaining. *See* Bargaining
  bluffing, 413
  currency, 70
  day. *See* Day trading
  fat fingers, 425–427
  flash crash, 415–416
  front running. *See* Front running
  hot potato volume, 426

Trading (*Continued*)
illegal insider, 404–411. *See also* Insider trading
information and, 171–172
late trading. *See* Late trading
market manipulation. *See* Market manipulation
payment for order flow, 423–424
Ponzi schemes and. *See* Ponzi schemes
risk in forward exchange markets, 68
rogue traders. *See* Rogue traders
rogue trading, 427–432
scale of, 26–27
securities. *See* Securities trading
spoofing. *See* Spoofing
textbook, algo programs for, 427
Trading arcades, 24
Trading floor broker, 39, 43
Trading platforms, 20–21
software-based, 20
web-based, 20
Trading strategies and performance, evaluation of, 297
implementation shortfall. *See* Implementation shortfall
investment portfolio performance. *See* Portfolio performance, evaluation
market timing vs. selection, portfolios and, 303–306
dummy variable approach, 305–306
quadratic variable approach, 303–305
trade evaluation, 306–308
VaR model. *See* Value at risk (VaR) model
VWAP. *See* Volume-weighted average price (VWAP)
Transaction size
perfect foresight half spread and, 110, 111*t*, 112*t*
slippage and, 110, 111*t*, 112*t*
Transactions exposure, managing, 255–256
Transcontinental Reality Investors Inc., 331
Transfer agent, advisor as, 95
Transitivity, axioms, 324
Transparent markets, 5
Treasury bills (T-Bills), 55
Treasury bonds (T-Bonds), 56
Treasury notes (T-Notes), 56
Treasury securities, 3
Treynor ratios, 300–301
Triangular arbitrage, 248–249, 249*t*
Trivia Quiz, 339, 340*t*
Trust indenture, 141
The Trust Indenture Act of 1939, 141
Tversky, Amos, 325
12b-1 fee, 104
Two-dollar broker. *See* Independent broker

U

U.S. Federal Reserve System, 207
U.S. Treasury securities and markets, 55–56
UBID, 208
UBS, 8, 19, 422
UITs. *See* Unit investment trust issues (UITs)
Ultra ETFs (bull funds), 105
Ultra S&P 500 ProShares, 105
Uniform Securities Act, 137
Unit investment trust issues (UITs), 45
Unlisted trading privileges (UTP), 84, 143, 148
Unregistered investment companies, 106–108
bank trusts and private banking, 107
hedge funds, 107–108
pension funds, 106–107
private equity, 107
Upstairs markets
crossing networks and, 53–55
internalization, 54–55
OTC, 45–46
USG Corporation, 378
Utility of wealth, 323*f*

V

VA. *See* Veterans Administration (VA)
Value at risk (VaR) model, 311–315
equity portfolio, illustration, 312, 314*t*
Value line investment surveys, 376
Vanderbilt, Commodore Cornelius, 418
Vanderbilt v. Drew, 418
Vanguard, 103
VaR model. *See* Value at risk (VaR) model
Variance, 197–201, 215, 221–224, 439–440
risk, 198–204
Vasicek, 247–248
Vector, 446
Vega, 278
Vermont Teddy Bear, 157
Veterans Administration (VA), 57
Vickrey auction, 9
Volatility(ies)
historical indicators, 199–201, 200*t*
implied. *See* Implied volatilities
semistrong form efficiency and, 378–379
Volcker Rule, 149–150, 158–159
Volkswagen, 419
Volume-weighted average price (VWAP), 54, 307–309
calculation, 307
price and volume for, 308*t*
illustration, 308
trade evaluation and, 306–309
Von Neumann, John, 324–325
VWAP. *See* Volume-weighted average price (VWAP)

W

Waddell & Reed, 416
Wall Street Journal, 121, 328, 374, 376
Walrasian auction, 8
Walrasian markets, 177
Wash sales, 138, 420
Washington Post Co., 425
Waskal, Sam, 407
Weak form efficiency, 360–370
  filter rules and market overreaction, 363–364
  IPO anomaly, 367
  January effect. *See* January effect
  moving average techniques, 364–365
  price sequences, momentum and mean reversion, 361–362
  relative strength rules, 363
  small firm and price/earnings effects, 366–367
  sports betting markets, 369
Weather, investment returns and, 336–337
Web-based trading platforms, 20
Who Wants To Be a Millionaire?, 346
Width/tightness, market, 15
Wiener processes, 196, 198, 272, 291–292, 294
Window dressing, January effect and, 366
Winner's curse problem, 9, 11
WorldCom, 61, 147

X

Xpedior, 208

Y

Yield curve, 235–236, 239–240, 247–248
  bootstrapping, 241–244, 242*t*, 243*f*
Yield to call, 230–231
  vs. yield to maturity, 230–231
Yield to maturity, 230
  vs. yield to call, 230–231

Z

Zero-coupon issues, 56